REPORT RESUMES

ED 003 456 24

TEACHING SCIENTIFIC THINKING AT THE HIGH SCHOOL LEVEL.
BY- SHOCKLEY, WILLIAM MCDONALD, FREDRICK J.
STANFORD UNIV., CALIF., SCHOOL OF EDUCATION
REPORT NUMBER CRP-S-090 PUB DATE OCT 64
CONTRACT OEC-4-10-216
EDRS PRICE MF-$0.27 HC-$5.44 136P.

DESCRIPTORS- JUNIOR HIGH SCHOOLS, *MATHEMATICS INSTRUCTION, *SCIENCE EDUCATION, SCIENCE INSTRUCTION, *SCIENTIFIC METHODOLOGY, *TEACHING TECHNIQUES, TEAM TEACHING, SUMMER PROGRAMS, COGNITIVE DEVELOPMENT, STANFORD, CALIFORNIA

THE PURPOSE OF THE STUDY WAS TO TEST (1) THE EFFECTIVENESS OF A WAY OF THINKING ABOUT PROBLEMS AND (2) THE EFFECTS OF INTRODUCING JUNIOR HIGH SCHOOL STUDENTS TO THE LOGIC OF SCIENTIFIC THINKING. THE STUDY HAD TWO OBJECTIVES (1) TO TEACH THE PROCESS OF SEARCH FOR AN ANSWER AND (2) TO TEACH THE LOGICAL STRUCTURE OF IMPORTANT PRINCIPLES. TEACHING SESSIONS HAD TWO MAJOR PARTS (1) TEACHING A SET OF LOGICAL GAMES AND (2) CONDUCTING EXPERIMENTS TO DERIVE A SCIENTIFIC PRINCIPLE. THE STUDY WAS CONDUCTED IN TWO PARTS--(1) A 15-WEEK SESSION CONDUCTED BY 3 TEACHERS IN A TEAM TEACHING ARRANGEMENT AND (2) A SMALLER SUMMER SESSION CONDUCTED BY 1 TEACHER USING AN INNOVATION IN TEACHING METHOD. WHILE THE RESULTS MUST BE CONSIDERED SUGGESTIVE RATHER THAN CONCLUSIVE, SIGNIFICANT COGNITIVE DEVELOPMENT WAS NOTED ON VARIOUS MEASURING DEVICES. THE LACK OF UNIFORM RESULTS WAS IN PART ASCRIBED TO DIFFICULTIES IN ESTABLISHING A TIGHT RESEARCH DESIGN. HOWEVER, THE STUDY HAS SHOWN THAT GIVEN APPROPRIATE CONDITIONS, REFINEMENT OF PROCEDURES, AND SENSITIVE TESTING DEVICES, A SIGNIFICANT DEMONSTRATION OF THE EFFECTS OF THE PROGRAM ALMOST CERTAINLY COULD BE ACHIEVED. (PM)

ED003456

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

TEACHING SCIENTIFIC THINKING AT THE HIGH SCHOOL LEVEL

by

William Shockley
School of Engineering

and

Frederick J. McDonald
School of Education

FFICIAL RESEARCH FILE COPY

October 1964

FINAL REPORT

Prepared under
Department of Health, Education, and Welfare
Office of Education
Contract No. OE 4-10-216
Project S-090

School of Engineering
School of Education
Stanford University Stanford, California

TEACHING SCIENTIFIC THINKING AT THE HIGH SCHOOL LEVEL

by

William Shockley
School of Engineering

and

Frederick J. McDonald
School of Education

October 1964

FINAL REPORT

Prepared under

Department of Health, Education, and Welfare
Office of Education
Contract No. OE 4-10-216
Project S-090

School of Engineering
School of Education
Stanford University Stanford, California

FOREWORD

This report in itself is an example of the research process that constitutes one of the main items of the experiment. The investigation has ended while the "hunch-familiarity" interaction, as discussed in Fig. 1 of Section 1, was still developing towards the "result" stage. This is particularly apparent in Section 3 where the analytical aspects of the investigation are described in part in historical sequence. In fact one of the most definitive results is reported in Subsection H of Section 3, which was added after the remainder of the section was typed.

Even after the report was being finally processed for reproduction, the analysis of results on "abstract reasoning" gave very encouraging indications that the new teaching approach had actually produced significant improvements in mental functioning of the ninth grade high ability students. This is discussed particularly in connection with Table 11 on page III.19. Here and also in Subsection H on page III.26 there is evidence that the mental growth measured by the "abstract reasoning" has been moved nearly six months ahead by modifying 15 weeks of the general science curriculum. Some even more striking results in a more limited area are discussed in Subsection IIIB page III.3.

Evidence that the new type of teaching employed was beneficial to the teachers temselves is found in Appendix 4 written by two teachers in the 15 week ninth grade programs and also in Appendix 6. The particular discussion presented in Appendix 6 page A6.22 shows how one teacher gained greater insight into the internal coherence of science and transferred knowledge learned in one area to other areas.

As the reader will appreciate, the report is based on experiences occurring under a variety of circumstances. Some of these had no original connection with the program and were exploited as opportunity offered.

The findings of the project directors are presented in Sections 1 to 4. In addition a mathematical analysis by Dr. Shockley is given in Appendix 3.

Samples of materials used are given in Appendices 1 and 2. Appraisals by teachers, by the Research Department of the Palo Alto Unified School District and by a Vice-Principal are given in Appendices 4 to 7.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to Dr. E. G. Begle for his encouragement to experiment with this material at the ninth grade, to Doctors George Polya, I. J. Quillen and P. D. Hurd for their interest and encouragement, to Dr. L. E. Moses for helpful discussions concerning statistical problems, and to the teachers and administrators in the Palo Alto Unified School District and in the Ohio Science Education Association. We would also like to acknowledge the contributions of Charles S. Whitmer, Francis A. J. Ianni, and Mrs. Polly Greenberg, to the creation of the circumstances which led to carrying out this experiment.

CONTENTS

Page

I. Rationale . . . . I.1 to I.6

II. Methodology . . . . II.1 to II.7

III. Test Instruments, Statistics and Tests of Significance . . . . III.1 to III.26

- A. Introduction III.1
- B. The Checkerboard Test III.2
- C. The Relative Rank Index III.9
- D. Relative Rank Studies for DAT Abstract Reasoning and Verbal Reasoning III.14
- E. Application of the Relative Rank Index to Study of a Test Based on SMSG Geometry III.21
- F. The Logical Structure Test III.23
- G. Apparent Age Shift III.24
- H. Note Added in Final Draft III.26

IV. Conclusions . . . . IV.1 to IV.3

APPENDICES

Page

1. Experiments A1.1

2. Tests A2.1

3. Significance Test for the Relative Rank Index A3.1

4. An Evaluation of Dr. Shockley's Program by Two Teachers at Wilbur Junior High School A4.1

5. Evaluation of Science Questionnaires at Wilbur Junior High School A5.1

6. Report by Miss Humphrey on Cubberley Teaching Experiments Summer 1964 . . . . . A6.1

7. Observations of the Shockley Project for Teaching Scientific Thinking at the High School Level by George Millar, Vice-Principal, Cubberley High School A7.1

I. RATIONALE

The purpose of the studies reported here was to test the effectiveness of a way of thinking about problems and to test the effects of introducing junior high school students to the logic of scientific thinking.

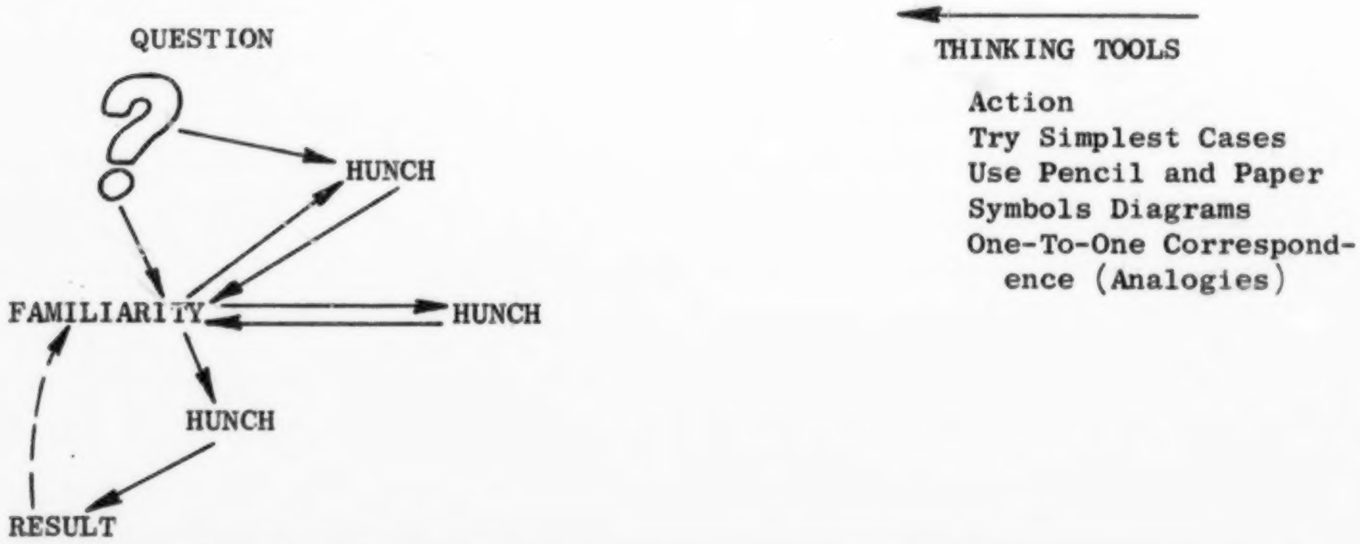
The search for general thinking tools is recurrent in the history of mankind. The object of this search has been to find general ways of thinking that enable man to solve many different kinds of problems, to generate ideas as diverse as poetry and scientific theory. The search has usually rested on a theory of mind or a theory of thinking. But even as anti-mentalist a philosopher as John Dewey wrote a book on thinking.

The recent history of education has revived interest in cognition. Bruner's speculation and research on such processes as discovery, intuitive and analytic thinking; the work of Torrance on creative thinking and that of Suchman on critical inquiry represent significant instances of the current and widespread interest in thinking processes.

This revival has occurred after a long period of disinterest among educators and psychologists in theories of general transfer. Thorndike's work on transfer at the turn of the century and his attack on faculty psychology and theories of mental discipline probably was the single most significant antecedent of this decline in interest. There followed a long period of educational research in which the major conclusion always seemed to be that transfer of thinking skills might occur but had to be taught for specifically. This view has prevailed until the recent interest in thinking processes which has been stimulated by the curricular innovations.

This project is in the spirit of the revival. It was stimulated originally by the observation that mere revision of content in science and mathematics courses did not seem to produce the massive change in thinking that some people seemed to expect. Even casual observation revealed that students did not intuit the logical structure of the new subjects, or even see the relation between concepts and principles learned in one topic to those learned in another. These observations are not offered as proof of the assertions made. They were the stimuli that prompted the ideas developed and tested in this project.

One aspect of these ideas is the general schemata for attacking any problem requiring analysis. This schemata might be called a "plan" for behaving in a problem situation. It summarizes the personal experience of one of the authors in research and invention and contains the principal features of several other summaries. It consists of two main parts: 1) The Familiarity-Hunch-Result search process employing Thinking Tools and 2) The Result expressed as "What is true, when it's true, and when it's not true." This material may be represented in a diagram as shown in Fig. 1.*


QUESTION
?
HUNCH
FAMILIARITY
HUNCH
HUNCH
RESULT
THINKING TOOLS
Action
Try Simplest Cases
Use Pencil and Paper
Symbols Diagrams
One-To-One Correspondence (Analogies)

WHAT IS TRUE ("Always true")
(A theorem, a principle or law, or an exact statement or equation)

WHEN IT IS TRUE ("no matter what")
(Range of application; examples)

WHEN NOT TRUE ("but not if")
(Area beyond range of application; examples)

Figure 1. Shockley's representation of the thought process of developing a principle.

The process of Fig. 1 can be illustrated clearly as discussed in Sec. 2 with the aid of carefully selected games and mathematical problems with underlying principles closely related to laws of science.

---

*This diagram presents essentially the same material as "Proposed important mental tools for scientific thinking at the high school level," W. Shockley, Science, 140, p. 384, 1963.

Figure 1 differs in an essential aspect from other outlines of the thinking process. To illustrate the relationships, three examples are quoted below:

> "Planning for Effective Science Teaching," T. G. Aylesworth, American Education Publications (1963):
>
> "John Dewey outlined this (method of critical thinking) in a series of steps:
>
> (A1) Identify problem
> (A2) Establish facts
> (A3) Formulate hypothesis
> (A4) Test hypotheses
> (A5) Evaluate results."
>
> "How to Solve It," G. Polya, Doubleday Anchor Books (1957).
>
> (P1) Understanding the Problem.
> What is the unknown? What are the data? What is the condition?
> (P2) Devising a plan.
> Do you know a related problem? Look at the unknown! Here is a problem related to yours and solved before. Could you use it?
> (P3) Carrying out the plan.
> (P4) Looking back.
> Can you check the result? Examine the solution.
>
> "A Technique for Producing Ideas," James Webb Young, Advertising Publications (1952):
>
> (Y1) The gathering of raw materials--both the materials of your immediate problem and the materials which come from a constant enrichment of your store of general knowledge.
> (Y2) The working over of these materials in your mind.
> (Y3) The incubating stage, where you let something beside the conscious mind do the work of synthesis.
> (Y4) The actual birth of the Idea--the "Eureka! I have it!" stage.
> (Y5) The final shaping of the idea to practical usefulness.

A description closely resembling Young's outline is given in Hadamard "Psychology of Invention in the Mathematical Field," Princeton University Press. In particular, Hadamard emphasizes the incubation stage (Y3) and its importance in many great discoveries. The close resemblence of the views coming from such different fields as mathematics and advertising (Young was founder in the Advertising Council, the creators of Smokey the Bear) suggests that certain features of the thought process of developing a new idea reflect characteristics of the mechanics of the functioning of the human brain and that this functioning or cognitive skill may be improved to a significant degree by teaching students the approaches based on the representation of Figure 1.

The essential feature that Fig. 1 emphasizes in contrast to the other outlines is commonly acknowledged by most creative researchers: the process of developing a principle or bringing order out of chaos is not always orderly. The development seldom follows the logical (A1) to (A5) sequence of Dewey. Sometimes the problem cannot be identified clearly until the "Result" has nearly been obtained. In Fig. 1 the question mark labelled "question" represents a wide range of possibilities including as one example, a specific word problem in ninth grade algebra and as another vague feeling that somehow the low current-carrier densities in semiconductors can be controlled to make a useful electronic amplifier. In a case like the latter clearly identifying the problem was the main research objective that lead to the transistor. In general "question" is anything which provodes the brain into trying to establish some better order or understanding.

The most important thinking tool is "action." This helps to store in the active part of the brain the relevant materials among which new relationships may be discovered; the result of this storing is called "familiarity."* Intuitive appreciation of new relationships are called "hunches." A wrong hunch may be valuable; it leads to more action and familiarity.

The "action," "familiarity," and "hunch" part of the representation corresponds to (Y1) (Y2) of Young, (P1) (P2) of Polya and to some degree (A2) (A3). The final sequence from "hunch" to "result" (not actually as simple as shown; there may be more steps back to "familiarity") represents (A4) (A5), (P3) (P4) and (Y5).

The incubation or (Y3) aspect, which is vital in inventing such new concepts as the benzene ring, the junction transistor, and (it is conjectured) Smokey the Bear, is probably a necessity at the Ph.D. candidate level whereas one can graduate from high school without learning it. It constitutes a somewhat higher state of sophistication than is represented in Fig. 1.

---

*The large variations of scientific and inventive ability can probably be explained in terms of differences in capacity to store and compare ideas as well as in maturation. W. Shockley, "On the statistics of individual variations of productivity in research laboratories," Proc. Inst. Radio Eng., 45, 279 (1963).

An example of a "result" used in the teaching experiment is a more exact statement of "What goes up must come down.":

> "It is always true that a solid object projected upwards on earth must return to the surface, no matter what it is made of, but not if it is accelerated to reach orbital or escape velocity."

If an appreciation of the disorderly aspect of discovery and of learning shown in Fig. 1 can be imparted, it may contribute to offsetting feelings of incompetence and distaste developed by students who are taught that the "only right way" is a logical sequence like (A1) to (A5). Such students may become "mathematics-haters." Some of these emotional aspects of the experiment are discussed more fully in Sec. 2.

To summarize the discussion above: One objective was teach the process of search for an answer and a second was to teach the logical structure of important principles.

The educational problem is to teach this set of concepts and their application in science. The research problem is to find ways of teaching these ideas that are effective.

Complex concepts of this kind are not easily taught. The temptation is simply to state them and to have students learn how to repeat them back in rote fashion. This instruction is obviously not likely to produce any depth of understanding or insight. In this project a set of simple, interesting games and problems was used as a way of helping students to grasp these concepts.

All of these games were two or more person games, and generally resemble the board games with which most people are familiar. The first game used had the following characteristics: (1) a general principle could be formulated which predicted who would win the game; (2) the assessment of who would win was not immediately obvious, nor was it apparent either that the winner had invariant characteristics. In these regards the principle of the game resembled a law of nature. Typically a game began with no awareness of who was going to win, proceeded to a phase in which a description of some kind could be given of the winner, and from there gradually to a formulation of the principle that explained who would win.

The first game taught required data-gathering, making hypotheses, testing them, and eventually formulating a principle. The principle or "law" of the game could be stated as some thing that was invariantly true given a set of conditions. In these ways identifying the logical structure of the game resembled establishing a scientific principle in keeping with Figure 1.

The game discussed and other games in which the outcome depended on understanding and applying a principle were introduced to students as something interesting that they might like to learn and to play competitively with each other. Usually, the games were preceded by a brief description of the model of thinking explained above, and the games were described as requiring the kind of thinking one goes through in establishing scientific facts. The purpose of the games is to serve as device for illustrating the processes required in scientific thinking, the difficulties involved, and the logical structure of laws and principles under circumstances in which the student becomes actively involved.

The second problem is to apply this way of thinking to scientific problems. A series of experiments were devised to teach the concept of potential energy as an explanatory concept that underlies such principles as those of levers. The strategy in these experiments resembled that in the games. The student was presented with some data gathering to do, such as finding the center of gravity of an irregular object. The object was moved through several positions; in each case the student is asked to predict how the object will hang from a support—like predicting who will win the games. Then a principle is called for, to organize and express the invariance in the observed phenomena. Similar experiments are repeated in which the strategy is carried out, and which extend the student's understanding of the scientific principle or concept he has learned.

The games serve the purpose of introducing the concepts and the strategies in an interesting way. The experiments extend and apply the ideas. The hoped for attainment is the acquisition of a way of thinking about problems, and an insight into the invariant character of scientific laws and the thinking processes involved in their development.

II. METHODOLOGY

The work described in this study was conducted in two parts. A fifteen week teaching session was held with one hundred and fifty ninth grade students at Wilbur Junior High School in Palo Alto. These classes were taught by three teachers in a team teaching arrangement. The classes were selected because they represented a spread of ability and because the teachers were interested in the study and willing to cooperate in its execution.

These teaching sessions had two major parts: (1) in the first the students were taught the logical games referred to above; (2) in the second, they conducted a number of experiments to derive a scientific principle.

This sequence of events was repeated in a teaching session during the following summer. The purpose of this second trial was to test the ideas of the study under more favorable conditions. A smaller class, thirty in number, and only one teacher was used. An innovation in the method was also tried at this time. Half of the group was originally taught by the teacher, and then this group was used to do some of the teaching to the second half.

Treatments In both of these studies* the set of games was used to teach the steps in thinking through to a problem solution, and the way in which to formulate a general principle or law. The set of experiments required similar thinking processes. The students gathered data which led to the derivation of a general principle.

The following is a description of one of the games used for teaching purposes, the "Take One to Five, Last Match Wins" game. This is a two-man, competitive game. The object of the game is to be the player who takes the last match (matches, paper clips, or similar objects are used as counters in the game; marks on a blackboard can be used equally well in a classroom demonstration). The players sit with a pile of matches of some arbitrarily chosen number such as twenty-five between them. Chance determines the first play. Each player in turn takes matches from the pile. He may take as few as one or as many as five, or any number in between, at each play. The player who removes the last match, leaving none for his opponent, wins.

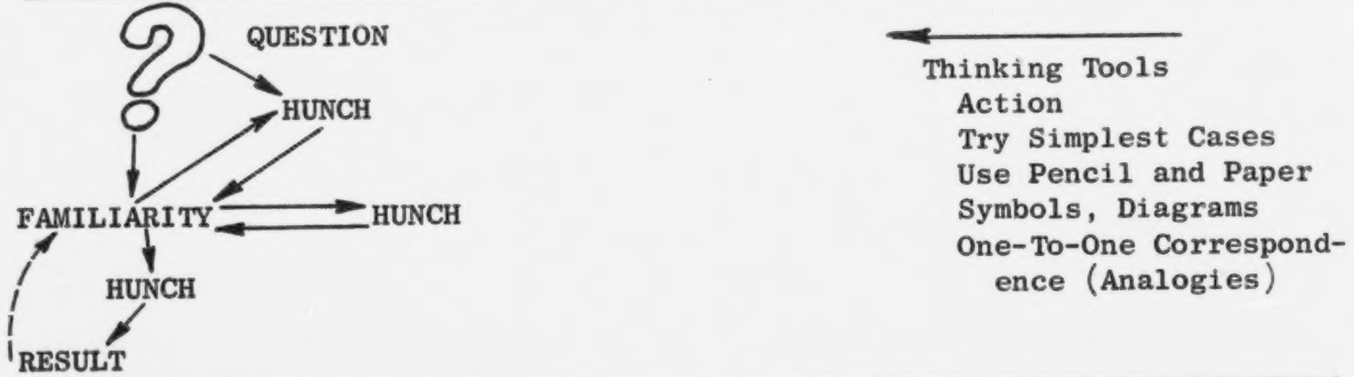
---

*Reports by teachers of both experiments are given in Appendixes to this report.

Students playing this game discover quickly that six is a winning number. If six matches are left, obviously the player who left them will win, no matter how many matches his opponent takes. His opponent can take no more than five, thus leaving one or more up to five which the winner can take. The next step for the student is to discover that leaving his opponent a group of twelve matches also guarantees winning, because after the next two plays it can be reduced to a group of six.

These experiences lead the student to the "hunch" that leaving groups which are multiples of six will insure a win if he plays properly. This hunch can be tested in repeated plays and verified. At this point the student has grasped the basic principle of the game at one level of understanding. Then the student is taught to state the principle in a form described in Figure 1, a form requiring him to state "what is true," "when it is true," and "when it is not true:"

> "It is always true the player who leaves a group of six or some multiple of six matches will win;
> "No matter what plays his opponent makes;
> "But not if he makes an error on a subsequent play."



WHAT IS TRUE ("always true")
(a theorem, a principle or law, an exact statement or equation)

WHEN IT IS TRUE ("no matter what")
(range of application; examples)

WHEN NOT TRUE ("but not if")
(area beyond range of application; examples)

Figure 1. Shockley's representation of the thought process of developing a principle.

At this simple level, the student has grasped and stated a basic relationship describing the game. However, a much deeper understanding of the principle underlying the game may be achieved. At this point the student is urged to state the principle in terms of the fundamental mathematical relations implicit in it. This more abstract and more general statement is as follows: "it is always true that a player can match his opponent's play so that the sum of their plays is six, no matter what play his opponent makes, but not if there are less than six matches when his opponent plays." In this form, the principle can be applied to a highly similar game, but one in which the player forced to take the last match loses. The student learns in this way the utility of general principles.

One of the advantages of using these games is that their subject matter is simple. Attention can be focussed on the logical structure of the problem, which is sufficiently difficult, without having to acquire many new concepts and facts. Also students can be taken quickly with the games through the steps in scientific investigation using a minimum of laboratory equipment.

Another characteristic of these games is that success in them depends on understanding. If a student grasps a principle, he not only knows the conditions under which he will win, but he also understands when he will lose and why. Obviously, in some of the games even Einstein could not win under certain conditions. The student learns the relation between the cognitive and emotional aspects of problem-solving.

They may also learn that in competitive situations understanding, thinking, reasoning gives them an advantage, and confidence that they will not always be "outsmarted." These competitive games are "real-life" situations—the opponents are really playing against each other to win, and winning brings recognition and satisfaction. The experience of thinking through a problem in a real-life situation like these may help the students learn that mathematical, logical, and scientific thinking may be used effectively in coping with life problems. A hoped for consequence would be that this learning would stimulate greater interest in the value of mathematics and science as ways of solving problems.

In summary the games are designed to teach what is required intellectually and emotionally, to solve a problem that involves using new logical structure. They are designed to demonstrate the general validity and utility of this kind of thinking.

These experiences with games are meant to introduce the student to a general way of thinking which he can apply in science. The question is then, how is a bridge made between the games and the scientific experiments?

The result of each game-playing experience is the formulation of a general principle. This principle is a statement of invariance under a set of conditions. The students are being taught to look for such invariances, because they are the devices for explaining, predicting, and describing what has been observed. The experiments to be described below are the means for gathering data which will lead to the discovery of a scientific principle, an invariance, that will explain these observations.

The first experiment (see Appendix 1) was called, "Experiment on Hanging a Rigid Body." The object of this experiment was to show that a principle could be discovered that explained the observations in the same way that a principle was discovered to account for winning in the games. The student was required to experiment with an irregularly shaped piece of plywood with a number of holes in it. He was to hang the board from these holes successively and to discover that a certain point on the board always hung below the point of support. The point was identified as the center of gravity and this concept defined. The student was then required to state the behavior of the board in respect to its center of gravity in terms of a general principle.

This experiment was followed by a number of others (see Appendix 1) in which the student made observations that would lead him to discover the conditions under which the principle about the center of gravity was true and not true. These principles and concepts were generalized in study of the laws of levers.

The study of levers led to establishing the concept of potential energy, and the study of levers as systems that conserve potential energy. With this concept grasped, the students performed experiments on the pendulum which is a system that conserves both potential and kinetic energy.

As can be seen, the students moved through a series of experiments which had the following characteristics: (1) the students always gathered the required data; (2) they were required to describe their observations in terms of a general principle; (3) they were required to state the conditions under which the principle was true and not true; (4) the principle and its associated concepts served as a basis for acquiring a broader generalization which explained more phenomena, e.g., levers in terms of masses located at centers of gravity, levers in terms of potential energy, pendulums in terms of potential and kinetic energy. The ultimate goal was to lead the students to an understanding of the law of conservation of energy as a comprehensive statement of invariance which could usefully organize a wide range of phenomena.

Subjects The students used in this study are described in greater detail in the section on results. The sample used obviously depended on availability and desire to participate of teachers. Ninth grade was chosen for a variety of reasons. General science is taught at this grade, and the material to be introduced seemed appropriate for a general science course. It was thought that some of the ways of thinking to be developed, and even some of the details of mathematical operations might serve the students in the following year when they studied geometry.

One of the purposes of the study was to find the age or grade level, or the degree and kind of cognitive development required to profit from this experience. No attempt was made to select a random sample of students from some carefully defined population. This procedure would be too refined for the stage of development of the experiment. For identical reasons, groups of comparison subjects were chosen who resembled the experimental group.

All subjects used in the study had known characteristics; that is, aptitude test data, grades, and teacher evaluations were available. Two groups of ninth graders, previously laned by ability were used as the study group. Other classes from the same and similar schools were used for comparison purposes. No attempt was made to randomly assign subjects to treatments, or to match groups by individuals. Differences among groups are described at various points in the reports.

Methods of Evaluation One of the purposes of the study was to develop methods of evaluation. It was immediately obvious that existing instruments did not sample the behavior being learned in the experimental sessions. Test items were developed and tried on an ad hoc basis. As a consequence, a number of novel and usable test items and methods were created, which subsequently will be developed into comprehensive testing procedures.

Test items, with few exceptions, were problems. They required time to solve, and consequently, an hour testing program might include only a few such problems. Another feature of the evaluation methodology was that a system of prompts and hints were introduced when it became apparent that problems presented real difficulty for many students. This procedure increased the number of correct solutions and close approximations to them.

The problems used seemingly met the requirements for valid evaluation of what was being taught. However, to achieve such validity the test procedures could not be cast in the form of typical achievement tests (with a few exceptions to be noted). That is, students learned on certain tasks as part of the experimental treatment and then performed on similar tasks for evaluation purposes. The word "tasks" is deliberately used in place of "test items." The research problem was to identify relevant evaluation tasks.

From among these tasks those were selected which appeared to be best related to the type of task used in the experiment. Consequently, some data were not analyzed that were gathered. The scope of the contract also did not justify elaborate analysis. The study produced a number of highly useful test problems and ideas for their further development.

Teachers participating in the study constructed achievement tests on the science content taught as part of the experiments performed. These teachers also prepared a set of questions which were used to assess the students subjective evaluation of their experience in the study (Appendix 5). Teachers also prepared a report of their evaluations and reactions (Appendixes 4, 6, 7).

Limitations of the Methodology This study was never intended to be a carefully controlled experiment with random assignment of subjects to treatments. It was a pilot study to test some "hunches" and to gain greater "familiarity" with the phenomena of concern. Every attempt was made to draw out of the data any significance that might be there. In short, the object was to answer the question, "Has anything happened, is it worth pursuing?"

All conclusions drawn should be evaluated by the design. These conclusions are couched as "as if" the design had met certain requirements. Differences among classes of students, differences among teachers and schools, variations in interest during testing are among some of the factors that could reasonably be expected to account for some of the differences among the groups (not all of which favored the experimental subjects), and to account for the lack of differences as well.

If these precautions are kept in mind, the following analysis may be seen for what it is, an attempt to tease out of the data any significance, any meaning that might lead to further useful experimentation.

III. TEST INSTRUMENTS, STATISTICS AND TESTS OF SIGNIFICANCE

A. INTRODUCTION

An objective of this contract is to evaluate the effect of the teaching on the ability to solve problems and to understand logical structure and laws of science. A number of tests were tried in the course of the program. The approach was itself the typical research approach represented in Fig. 1 of Section I that was being taught. The general conclusion is that suitable tests can be invented and perfected. The checkerboard test discussed in Sec. B below and the logical structure test of F illustrate approaches thought to be especially suitable.

In this section, we shall focus attention upon tests especially adapted to studying ability to do logical thinking and to transfer knowledge of logical structure to new areas. Tests having to do with evaluating the teaching of the laws of levers and the significance of conservation of energy, will be found discussed in the Appendices 4 and 6 written by the teachers.

In respect to teaching the understanding of the general features of a law or principles in the terms given under "Result" in Fig. 1 by stating what is true, when it's true, and when it's not true, a pre-test and a post-test showed obvious and pronounced improvement as was expected.

It was hoped that the longitudinal study test of SMSG mathematics would be available to help in this experiment. However, the 9th grade students for the 1963-64 school year had not been given this test. Probably the effect of the teaching of this experiment upon competance in mathematics is relatively small compared to the effect of individual classroom experience; this point was incidently investigated during this study and is reported in Subsection E below.

One of the procedures explored in the experimental tests was that of giving successive hints during the test so that students who might otherwise accomplish nothing would be brought along to some degree comparable to the best students. Their scores would recognize the time lag. This was accomplished in the checkerboard test in such a way that early performance is clearly distinguished from later performance.

One of the most provocative results of the study is in an indication that the training may have had a significant effect upon general cognitive skills as measured by the Differential Aptitude Test, particularly for abstract reasoning and also to a lesser degree for verbal reasoning. These results are consistent with results obtained from the checkerboard test and suggest that an increase in general cognitive skill corresponding to possibly 6 months of intellectual growth may have been accomplished, whereas tests on more specific problems having closely related logical structure amounted to between 2 and 3 years as discussed in respect to the checkerboard test particularly in regard to standing relative to the control group CO as discussed in connection with Table I below.

Some new statistical methods not generally available were required to compare relative progress of unmatched control experimental groups. This lead to the development of the relative rank index concept discussed in Subsection C which is very closely related to certain previously developed statistics.

B. THE CHECKERBOARD TEST

This test was intended to measure the ability to transfer the experience in finding general principles to new problems. The test was based upon a well-known problem which has been included in one of the SMSG textbooks. This problem will be found in the Appendix quotation of the text, but it will be worthwhile, for the benefit of the readers, to describe it here.

Consider an 8 x 8 checkerboard with shaded and plane squares. This board contains 64 squares. Imagine you have 31 dominos. Each domino may be placed on the board so as to cover two adjacent squares, the shape of the domino consisting itself of two squares each identical to a square of the checkerboard. The student is asked to analyze the problem of trying to place the dominos on the checkerboard so as to cover all squares but two, with these two uncovered squares lying at opposite extremes of a diagonal running from one corner to the diagonally opposite corner. (It may be remarked that this problem has posed considerable difficulties to quite competent mathematicians. It is necessary to have the correct hunch as to what is involved in solving the problem. If the

TABLE 1. CHECKERBOARD TEST GROUPS

EXPERIMENTAL GROUPS

9th Grade Lane A: EA = DoA + DrA + RA
9th Grade Lane B: EB = DoB + DrB + RB

CONTROL GROUPS

9th Grade Lane A: JA
9th Grade Lane B: JB
9th Grade Advanced Placement Biology: JF
11th and 12th Grade Ohio Honor Science Students: CO

problem is given to a mathematician as stated above, with mention made of the squares being shaded, he will be able to see it much more quickly than if it is given to him simply as an 8 x 8 square. In this latter case, an important fact is obscured, namely that the domino covers one square of each color; in fact, there are no colored squares to consider.) In terms of the colored squares, it can easily be seen that the total number of squares covered by the dominos always consists of equal numbers of plane and shaded. The problem as originally stated requires that the two squares left uncovered be of the same shading. Consequently, it requires that the dominos cover two more squares of one color than they do of the other color. This is obviously impossible.

The test incorporates an 8 x 8 board as well as a number of other shapes and sizes. This test had in common with the examples that had been taught chiefly the feature that a general underlying principle was involved. The problems previously taught did not involve checkerboards nor any case in which any single move always did equal things to two other things. Thus improved performance would suggest a transfer of a general method of attack.

Over 200 students were tested in all on this test. Many of these students had seen the checkerboard problem in SMSG courses. Of these, not more than 1 or 2 percent at most recognized this as a familiar problem. The form of the test gave many other examples than the 8 x 8 checkerboard, although the example was repeated in each test.

The results of the test indicate clearly that the students who had been subjected to the training outperformed the controls by such a wide margin that the null-hypothesis could be rejected at a significant level of less than 0.02. (The analysis will be discussed in somewhat greater detail below.)

In order to build a test to measure the ability to learn and to appreciate principles, some special features had to be incorporated. These were achieved by dividing the test into three parts. In Part I, the student was given problems which involved unshaded checkerboards, some of which could be solved and some not. Part way through this test, the student was asked to write down his general conclusions. This request was made by the teacher of the students and was emphasized. (Not more than one or two percent of the students tested proposed to add shading in this part.)

The first test papers were then collected and part two handed out. In Part II, the student was instructed to shade his checkerboard like a regular checkerboard and proceed as he had in Part I. He was again asked to write his conclusions. The striking results of this part of the test were that of the trained group in lane A (the faster lane as discussed below), approximately half of the class discovered the relevant principle once their attention was called to the shading. The percentage was very much smaller in the comparable control groups.

In Part III of the test, the principle was spelled out and the students were asked to apply it. This test gave little information about the students who had answered Part II correctly, and students of the lanes that were effective in Part II were not graded on Part III (because of limitations of the time in this small contract). A striking result was obtained for the slower students in lane B however, since these students responded significantly less well when asked to write statements about the principle in the third part of the test even though this had been explained to them.

The methods of analysis are shown in a series of tables of which Table 1 lists the groups that were compared. The experimental groups consisted of 9th grade students in Wilbur Junior High School in the Palo Alto Unified School District. Three separate classes in faster lane A were given the experimental teaching, and also three somewhat less fast classes in lane B. The students in the teaching experiment in lane A have a Differential Aptitude test in verbal and numerical reasoning (this is denoted by DAT/V+N and is roughly equivalent to an IQ rating) which puts them at the 87th percentile. The lane B group are somewhat lower on the average but are well above the 50th percentile.

The control groups were also 9th grade students in the Palo Alto Unified School District but from a different junior high school. Three classes were involved in the other junior high school. The most gifted of these is denoted as JF in Table I; these students are particularly competent in the field of science and have been placed in an advanced biology class. Another group is lane A and is denoted by JA in the table. This particular section of 30 students had approximately 17 at the 99th percentile on the DAT/V+N; the median percentile for this group was at the 99th percentile. The median for the advance placement biology class JF was at the 95th percentile.

In addition to these students, an opportunity occurred to test some science students in Ohio. These students were gathered for an occasion called "Project: Honoring the Ohio Science Student." They were typically 11th and 12th graders and represented the top 1 or 2 percent of the science students in their high schools.

The statistics of checkerboard test No. 2 which is described above is shown in Table II. In this particular test, the maximum possible score was 14 and the closest achievement to this maximum was made by one student of group JF who received 13. The three lane A experimental sections are treated as a group, the total number of students being 72. In the experimental group a score of 6 or below was made by 42 students or 58 percent while 30 students made 7 or above. On the other hand, for group JA a score higher than 6 was made by only 3 students so that 90 percent made a score of 6 or less. The difference of 42 percent between 90 and 58 is the largest difference between the cumulative percentages for the two groups. As is indicated on the table, the probability of this percentage occurring solely by chance on the basis of the "null-hypothesis H_o" that the group EA and JA were drawn from a common population can be evaluated by the Kolmogrov-Smirnoff test[*] and is found to be less than 0.02. Thus, the difference between groups EA and JA is significant at alpha = 0.02 level. This does not, of course, imply that this difference, if it is really representative of the difference in the two groups on this test, arose as a result of the training; however, the training was directed towards producing results of this sort.

TABLE 2. CHECKERBOARD TEST NO. 2

Cumulative numbers and percentages of students making scores equal or less than indicated scores (maximum possible 14).

Test Score	EA Group			JA Group			JF Group		
	No.	Cum.	Cum. Pct.	No.	Cum.	Cum. Pct.	No.	Cum.	Cum. Pct.
0	1	1		0	0		0	0	
1	1	2		1	1		0	0	
2	5	7		1	2		2	2	
3	5	12	17	5	7	23	2	4	1.5
4	11	23	32	9	16	53	7	11	40
5	10	33	46	6	22	73	5	16	62
6	9	42	58	5	27	90	2	18	69
7	5	47	65	1	28	93	2	20	77
8	10	57	79	0	28	93	2	22	85
9	2	59		0	28		1	23	
10	4	63		1	29		1	24	
11	1	64		0	29		0	24	
12	8	72		1	30		1	25	
13	0	72		0	30		1	26	
14	0	72		0	30		0	26	

	EA	JA	JF
Median Score	6	4	5
Median DAT/V+N	87	99	95

X^2 for EA vs JA by Kolmogrov-Smirnoff one-tail is 8.7 for df = 2, $P < 0.02$

The behavior of EA and JF is much more closely similar. This is not surprising in spite of JF's lower DAT/V+N median because performance shows that the JF group is gifted in sciences.

These differences between checkerboard 2 and the DAT/V+N scores may be given some interpretation in terms of performance on aptitude tests.

---

[*] See S. Siegel "Nonparametric Statistics," McGraw Hill, p. 135.

First, we note that group EA is performing better than JA on the checkerboard test although the median aptitude of EA is 10 percentile points below that of JA. If one had only the performance scores and wished to estimate aptitude, his best judgment, other things being equal, would be that EA was superior in aptitude. Since we know that EA in fact is not superior, the training effect may be looked at as an equivalent aptitude growth increment.

The median percentile ranks and equivalent raw scores for the respective ranks are given in Table III for 9th, 10th, and 11th grade students. (Male norms.) If the students in EA were to move in aptitude test scores in a normal growth increment, they would obtain a raw score of 66 for the 85th percentile two years later in the 11th grade. On Checkerboard 2, however, EA performed similarly to JF whose median percentile rank was 95 and whose raw score was 67.5. To obtain an equivalent raw score while remaining at the 85th percentile would require one year of growth.

TABLE 3. DAT/V+N RAW SCORES AND PERCENTILES

Percentile	RAW SCORES		
	9th Grade	10th Grade	11th Grade
99	66+	73.5	83+
95	58.5	67.5	74.5
85	49.5	58.5	65.5

Insofar as checkerboard 2 measures a performance capability generally similar to the verbal reasoning and abstract reasoning of the DAT tests, the data of Table II suggest that an effect has been produced, by the teaching, equivalent to about 18 months of normal growth. It should be understood, of course, that the authors have not reached this conclusion from the Differential Aptitude Test data, but that this conclusion is a quantitative expression of the shift in performance as measured by Checkerboard 2.

A significant result at the alpha = 0.025 level was also obtained in a test in the results of the Checkerboard 3. In this case, only one question was used for the comparison. This was the question, mentioned earlier, in which the students were asked to write down what principles they were using. This question was written on the text of the test and, in addition, it was called for by the teacher who was administering the test. In this particular case, the same teacher administered the test to the control group JB as to his own class RB. Table IV shows the striking result that only 2 of his own students failed to write anything on this question when requested to do so, whereas nearly 50 percent of the control group left the question blank. This sets a level of significance which can be evaluated by the Kolmogrov-Smirnoff test, alpha = 0.025 level. It may also be noted that approximately twice as many students gave good answers on the test for the experimental group as for the control group.

TABLE 4. CHECKERBOARD TEST NO. 3

(Test based on question asking for statement of principle.)

	RB	JB	
Blank	2 (9 percent)	13 (48 percent)	D = 39 percent
Wrong	12	9	$D(22 \times 27)^{\frac{1}{2}}/(22 + 27)^{\frac{1}{2}}$
Poor	1	2	= 1.36 (Kolmogrov-
Good	7	3	Smirnoff-One-Tailed
	22	27	alpha = 0.025)

It should, of course, be mentioned that this is a very small sample and that personal and emotional factors such as the administration of the test by their own teacher in one case and by a stranger in the other case may have distorted the results.

The comparisons with group CO showed that it did not differ significantly from the EA group, the correspondence being in fact very close. This fact suggests that the lane A students performed on this test at a level comparable to Honor Science students who represent top percentiles in Ohio of better than 95 so far as science performance is concerned. From this point of view it suggests again that the training given EA had the effect of between 2 and 3 years of normal growth.

C. THE RELATIVE RANK INDEX

At the end of the term in which the experimental teaching was done, attempts were made to measure effects on general ability to deal with new problems involving logical thinking. In particular, the ambitious project was proposed of using the standardized differential aptitude tests for abstract reasoning and verbal reasoning. It was obviously ambitious to expect any improvement in these since the material taught was not specifically related and only the most general aspects of critical thinking would be expected to transfer. In addition, an experiment was carried out in which the 9th grade students were given copies of the 10th grade SMSG Geometry book which they would study in the following year. Chapter 2 of this book deals with theory of sets and with certain basic features of numbers as represented by points on a line. The students were asked to work as many problems as they could in this Chapter 2 during a one hour period. This equally ambitious project was intended to see if any improvement to cope with homework had been produced.

The results from these tests showed, on first inspection, little that could be regarded as a significant effect. The experimental students differed in both number and aptitude as measured by the DAT. This posed statistical problems for which there do not appear to be well organized statistical methods. In the course of endeavouring to see if any significant difference arose, a "relative rank index" was invented. This index is nonparametric. It appears to afford a sensitive measure for determining whether there has been a change in competence between two groups provided there is a large amount of overlap between the initial competences of the groups. It was found that the relative rank index is closely related to a quantity known as the Mann-Whitney U and to the sum of ranks used in the Freidman 2-way analysis of variance. It also relates closely to other sums of ranks used in nonparametric statistics. However, its use as a measure of relative standing of two groups does not appear to have received much attention.

The relative rank index can be most directly understood in terms of concrete examples. Table V shows hypothetical examples for the case of two groups, one consisting of $n_1 = 3$ individuals drawn from a population

A, and $n_2 = 4$ drawn from a population B. It is assumed in Table V that on an initial test, called Test 1, the scores of all of the 7 students were different and that all of the A's obtained lower scores than all of the B's. For this situation, R_{12} the relative rank index of the n_1 sample in respect to the n_2 sample has its lowest possible value of -1. Basic to the concept of the relative rank index is the rank point. If, as shown in Test 2, the highest A student catches up and ties the lowest B student, then group A scores 0.5 rank points shown as U_{12} in the table; and group B, which originally had $U_{21} = 12$ rank points, looses half a rank point. If in Test 3 the student passes the B student, then group A is credited with 1 rank point and U_{12} and U_{21} change by +1 and -1 compared to Test 1. If finally, as shown in Test 6, all of the A students pass all of the B students, then U_{12} is obtained by giving each A student 4 rank points for the 4 B students he has passed, so that the total number of rank points for the A group is 12. Evidently in general the maximum for U_{12} is n_1n_2 and the same is true for U_{21}.

The relative rank index is defined in such a way that, as the A group gathers more rank points, its relative rank index increases from -1 when all of the A's are below the B's to +1 when all of the A's are above. This leads to the formulae at the bottom of Table V for R_{12} in terms of the total number of rank points U_{12} gained by the A group as well as to other relationships which are helpful in making and checking calculations.

The relationships shown in Table II may be readily verified. The following remarks may be helpful: the quantities U'_{12} give no rank points for ties to either side. Since when tied rank points occur they always occur equally for the A group and the B group, this does not affect the difference $U_{12}-U_{21}$ between the rank points obtained by group A and the rank points obtained by group B; is also shown below the table. Still another quantity which is closely related to the relative rank is the sum of ranks for all the members of the groups. Against each position in the table is seen the rank of the student or students who are tied. In the case of a tie, the average rank is given to each tied student. When these ranks are summed for each group, the sums r_1 and r_2 are obtained. These sums of ranks increase from a minimum value to a maximum value linearly just as do U_{12} and R_{12}.

TABLE 5. EXAMPLES OF RELATIVE RANKING INDEX VALUES

$R_{12} = -R_{21}$	Test 1 -1	Test 2 -0.92	Test 3 -0.83	Test 4 -0.583	Test 5 0.0	Test 6 1.0
Highest	B 7	B 7	B 7	B 7	B 7	A 7
	B 6	B 6	B 6		A 6	(AAA) 6
	B 5	B 5	B 5	(ABB) 5	A 5	
	B 4	(AB) 3.5	A 4		B 4	B 4
	A 3		B 3	(AB) 2.5	B 3	(BB) 2.5
	A 2	A 2	A 2		B 2	
Lowest	A 1	A 1	A 1	A 1	A 1	B 1
U_{12}	0	0.5	1	2.5	6	12
U_{21}	12	11.5	11	9.5	6	0
U'_{12}	0	0	1	1	6	12
U'_{21}	12	11	11	8	6	0
T	0	1	0	3	0	0
$r_1 = \sum r_{1i}$	6	6.5	7	8.5	12	18
$r_2 = \sum r_{2i}$	22	21.5	21	19.5	16	10

$$r_1 r_2 = 3 \times 4 = U_{12} + U_{21} = U'_{12} + U'_{21} + T$$

$$U_{12} - U_{21} = U'_{12} - U'_{12} = 2\,U_{12} - n_1 n_2$$

$$r_1 = U_{12} + n_1(n_1 + 1)/2, \quad r_2 = U_{21} + n_2(n_2 + 1)/2$$

$$R_{12} = -R_{21} = (U_{12} - U_{21})/n_1 n_2 = (2U_{12}/n_1 n_2) - 1$$

A particularly interesting feature of the relative rank index is that, if one of the groups is a control group which may be regarded as representing an entire population, then the U_{12} value gives the average percentile position of group 1 in respect to this population.* This conclusion may be reached by considering the rank points added by each individual member of the n_1 sample of Table V, as shown in Table VI. These individual contributions like $U_{12,3}$ have an additional subscript indicating the individual member of the A population who contributed the rank points.

---

*This feature has been discussed by Irwin D. J. Bross who introduces the "ridit": "reference to identified population" which is identical with $p_{12,i}$ of Table VI. Biometrics 14, 18 (1958).

Evidently, an individual A_i of the n_1 members of population A who attains $U_{12,i}$ rank points over the n_2 members of population B falls at a percentile position of

$$p_{12,i} = 100\, U_{12,i} / n_2$$

in respect to population B. Evidently,

$$U_{12} = \sum_{i=1}^{n_1} U_{12,i}$$

The average percentile position of the n_1 sample, obtained by simply averaging the n_1 values of the $p_{12,i}$, is therefore given by

$$\begin{aligned} p_{12} &= (\Sigma\, p_{12,i})/n_1 = 100\, U_{12} / n_1 \\ &= 50(R_{12} + 1) \end{aligned}$$

where the previous expression for R_{12} in terms of $2U_{12}/n_1 n_2$ has been used to eliminate U_{12}.

The fact that p_{12} is simply related to R_{12} shows that, if the sample of n_2 members of population B represents accurately some large reference population, then R_{12} can be used to calculate the mean or average percentile position of the n_1 individuals from population A on the percentile scale of population B. This average percentile position of the sample from population A varies from the lowest or 0 average percentile for the $R_{12} = -1$ to the 50th percentile when $R_{12} = 0$ up to the top or 100th percentile when $R_{12} = +1$.

This behavior of R_{12} is one of the features that makes it particularly appealing in comparing two populations. Another advantageous feature is that the relationship of R_{12} to U_{12} permits use to be made of the Mann-Whitney U test to determine the significance of observed differences of R_{12} values from the expected value of 0 for the null-hypothesis that populations A and B are the same.

In analyzing data in this experiment, increases of p_{12} of experimental groups A in respect to control groups B imply an upward shift of the experimental groups in respect to the population as a whole. A significant test is discussed in Appendix 3.

As will be discussed in the following sections, there are provocative pieces of data indicating that a significant improvement may have been achieved.

TABLE 6. AVERAGE PERCENTILE POSITIONS OF A SAMPLE IN RESPECT TO A REFERENCE POPULATION

Test	1	2	3	4	5	6
$U_{12,3};P_{12,3}$	0;0	.5;12.5	1;25	2;50	3;75	4;100
$U_{12,2};P_{12,2}$	0;0	0;0	0;0	2;50	3;75	4;100
$U_{12,1};P_{12,1}$	0;0	0;0	0;0	0;0	0;0	4;100
U_{12}	0	0.5	1.0	4.0	6	12
P_{12}	0	4.1	8.3	33	50	100

D. RELATIVE RANK STUDIES FOR DAT FOR ABSTRACT REASONING AND VERBAL REASONING

As discussed in Sec. C, the relative rank index arose in the course of this study in an attempt to find a sensitive measure which would show if relative progress of the experimental group could be observed in respect to the control group. In keeping with the discussion of Fig. 1, the evaluative research of this experiment followed the "familiarity, hunch and result" sequence that constitute the subject matter for the teaching experiment. As the work progressed, additional ideas developed and finally a result in terms of a significance test for the null-hypothesis was developed.

We shall follow the sequence of carrying out these investigations by presenting tables in the order that they were developed.

Table 7 shows the experimental and control groups involved in a study (some additional groups from another school were also tested; however, time has not permitted analyzing their data to see how it fits in the pattern. Nothing has occurred to suggest that any change in tentative conclusions would be produced by including this other data). The experimental groups in lane A are the same as those discussed earlier. The lane B experimental group has been dealt with only as one unit and no attempt to deal with individual classes has been made. The control groups are all indicated with the letter C, and in most cases the lane B control group has been omitted when comparing the controls with the lane A group.

TABLE 7. GROUPS IN RELATIVE RANKING INDEX STUDIES

Experimental Groups

9th Grade Lane A:	EA = DoA + DrA + RA
9th Grade Lane B:	EB

Control Groups

9th Grade Advanced Placement:	CF	C
9th Grade Lane A:	CA	
9th Grade Lane B:	CB	

Table 8 shows the first comparison between the groups. In this group, each student was assigned a score in the first test which was simply his percentile ranking on the differential aptitude test for abstract reasoning as given in the fall of 1963. The possible percentiles are 99, 97, 95, 90, 85, etc., at 5 percent intervals. On the basis of these percentiles used as scores, relative ranking indices and percent were given for the 68 students in experimental group EA versus the 3 control classes containing 63 students. As will be seen, this relative rank index expressed in percent increased from -1 to +13 on the 2 tests. A much smaller improvement was observed for the lane B group.

TABLE 8. RELATIVE RANKING INDICES IN PERCENT FOR DAT/AR FALL 1963 AND JUNE 1964

	Fall 1963	June 1964
EA (68) vs CF + CA + CB (63)	- 1	+13
EB (63) vs CF + CA + CB (63)	-20	-18

Some aspects of the comparison were observed which could distort the statistics. A relatively large fraction of the control group consisted of students who ranked in the 99th percentile in both tests. When the percentile position is used as a score, these students cannot improve their position in a ranking table. Consequently, any experimental students who rose from lower ranks would each receive half rank-points at the expense of students who they tied in the top 99th percentile.

It was decided to investigate this aspect more thoroughly, and in order to obtain additional information attention was shifted from the differential aptitude test for abstract reasoning to that for verbal reasoning. . Table 9 shows the results of these calculations. In this case, a total of 6 relative rank indexes were computed, for each two separate dates; i.e. each of the 3 lane A classes was compared with the two control classes. The results for the fall of 1963 are shown at the upper left-hand portion of the table. If we consider, for example, experimental group DoA containing 22 students, we see that its relative ranking on a percentage basis with the fast biology group CF containing 29 students was -29 percent in the fall of 1963. A corresponding ranking based on the test of June 1964 gave them a relative rank of -7 percent.

Thus, the change in rank as shown on the table was +22 percent indicating that the experimental group increased its ability to cope with these verbal reasoning problems faster than the control group.

TABLE 9. RELATIVE RANKING INDICES R_{ec} IN PERCENT FOR DAT/VR
FALL 1963 AND JUNE 1964 (R_{ec} IN PCT.)

FOR ALL 1963 PERCENTILES						
	Fall 1963		June 1964		Change	
	CF	CA	CF	CA	CF	CA
	(29)	(20)	(29)	(20)	(29)	(20)
DoA (22)	-29	+27	- 7	+26	+22	- 1
DrA (22)	-31	+26	+14	+18	+45	- 8
RA (23)	-30	+16	-22	+13	+8	- 2
FOR 95-65 INCLUSIVE 1963 PERCENTILES						
	Fall 1963		June 1964		Change	
	CF	CA	CF	CA	CF	CA
	(14)	(11)	(14)	(11)	(14)	(11)
DoA (15)	-26	+ 2	- 2	0	+24	- 2
DrA (16)	- 3	+20	+ 3	+ 8	+ 6	-12
RA (12)	-10	+16	+10	+21	+20	+ 5

The upper portion of the table shows 6 such comparisons, and it is seen that in all cases the lane A group progressed in respect to the fast control group CF.

The same change was not observed when the lane A control group was studied. This control group appeared to move somewhat ahead of the experimental group. In both cases, it is possible that a major effect was due to crowding of the students in the top percentile.

In an attempt to overcome this difficulty, the calculations were repeated omitting students who were initially in the 97th and 99th percentile. This means that students who could not rise to a higher percentile were omitted from the test so that a comparison would be made between students who had two percentile ranges into which to move. In this case, it is observed that 4 of the 6 groups had improvements in their relative ranking in respect to the control groups. However, the data does not appear to be strikingly significant.

It was realized that a test for significance would be required in dealing with such data. Since the groups are not matched, either in numbers or in initial capability, this posed a problem for which ready answers did not appear to be available in statistical references.

It was desirable to invent a means of testing the null-hypothesis that the experimental training had no influence on the behavior of the students and their improvement between the fall of 1963 and June of 1964. In searching for such a test of significance on null-hypothesis, no consideration was given to additional factors such as the possibility that students might be influenced by special teaching situations in other classes, such as mathematics classes. We shall return to this point in the next section in respect to a test based on SMSG geometry.

The method finally used for testing significance consisted of treating students in an original percentile position in the fall of 1963 as being comparable. Table 10 shows an analysis based on verbal reasoning. As will be seen, there were 15 students of the experimental group in the 90th percentile for verbal reasoning in the fall of 1963, and 4 students of the control groups. These students, as well as all of the others in the table, were compared in rank on the basis of their raw scores on the 1964 verbal reasoning test. The 90th percentile group thus achieved some 19 rankings, some of which were duplicates on the 110 final raw score ranks. The null-hypothesis, as is discussed more fully in an Appendix 3 consisted of calculating how many rank-points these students would have been expected to collect for experimental group EA, and observing that the actual number exceeded this expectation by 15 rank points as shown in the table. Furthermore, the numbers of individual rank-points involved are sufficiently large that, if they were selected at random for the 19 students of the 90th percentile, their distribution should be well represented by a normal distribution. Consequently, it is possible to calculate a mean squared deviation expected on the basis of the null-hypothesis for the 19 students in the 90th percentile. As is shown, this total number of rank-points RP_{EA} would be expected on a null-hypothesis to have a mean squared deviation of 1680. The null-hypothesis leads to independent statistical fluctuations for each of the percentiles.

TABLE 10. NORMAL DISTRIBUTION SIGNIFICANCE TEST OF DAT/VR CHANGES FROM FALL 1963 PERCENTILES TO JUNE 1964 RAW-SCORE RANKINGS

DAT/VR Fall 1963	Students EA	Students CF+CA	RP_{EA} (EA Rank Points Excess Over H_o)	Sigma Squared for RP_{EA}
99	10	10	-25	1560
97	9	9	+43	1820
95	7	7	+22	650
90	15	4	+15	1830
85	7	7	+22	1680
80	5	3	- 6	455
75	--	--	--	--
70	2	3	0	208
65	4	3	- 6	140
60	3	2	- 6	20
Totals	62	48	59	8363

$z = 59/(8363)^{\frac{1}{2}} = 0.625$

One-tail $P < 0.26$

Omitting 60-70, $z = 0.9$; $P < 0.18$

Increase of $R_{ec} = 59/62x48 = 0.02$

This proceedure permits comparing the total excess rank-points or 59 above the null-hypothesis with the number to be expected for a normal distribution *represented by the independent action* of the 9th percentile groups of Table 10. As is shown in the table, this leads to a one-tailed level of significance probability of approximately 0.26.

The numbers of students in the 60th, 65th, and 70th percentile is relatively small. A discussion with the counselor who placed these students in the sections shows that he considered that the percentile rankings of 60, 65 and 70 were not representative of the other performance expected of these students. Omitting these three percentiles leads to the lower one-tailed probability of 0.18.

The results of this study are provocative and suggest that a gain in ability to do verbal reasoning may have been achieved by the experimental program, but that the significance with which such an assertion can be made is very low.

With the statistical methods developed for Table 11, the abstract reasoning was again studied with the results shown in Table 11. In this case, the increase above the null-hypothesis H_o is seen to be very much larger, and the probability that so large a result would have come about from the null-hypothesis alone is less than 0.021. The increase in relative rank on the basis of an excess of 192 rank-points above a zero corresponds to an increase of relative rank of 0.13 which is quite consistent with the first results obtained for the lane A students, as shown in Table 11.*

(Corresponding studies have not been carried out for the lane B students due to limitations of time and manpower on this small exploratory contract.)

TABLE 11. NORMAL DISTRIBUTION SIGNIFICANCE TEST OF DAT/AR CHANGES FROM FALL 1963 PERCENTILES TO JUNE 1964 RAW-SCORE RANKINGS

Fall 1963	Students EA	Students CF+CA	RP_{EA} (EA Rank Points Excess Over H_o)	Sigma Squared for RP_{EA}
99	7	8	+ 37	2340
97	6	6	+ 1	1050
95	8	4	- 9	1430
90	7	2	- 2	480
85	12	5	+144	3040
80	7	3	+ 11	552
75	1	3	+ 10	36
Totals	48	31	192	8928

$z = 192/(8928)^{\frac{1}{2}} = 2.04$

One-tail $P < 0.021$

Increase in $R_{ec} = 192/48x31 = 0.13$

The results of the study are suggestive that the type of teaching given in connection with the objectives of Fig. 1 and the related scientific subject matter pertaining to the laws of levers and the application of mathematics to understand science may have had a general effect on cognitive skill as measured by the DAT/DR and DAT/AR. In considering this possible conclusion, several comments are worth making. The training given in the teaching was not directed specifically towards these tests. In fact, no plan was made to give these tests until about one week before the end of the term, when the "hunch" developed that this might be a worthwhile experiment. The subject matter of these tests is quite different from the subject matter being taught. An example of the verbal reasoning question will indicate its nature:

---

* See also subsection IIIH.

. . . is to water as eat is to . . .

1. Continue 2. Drink 3. Rain 4. First
A. Two B. Fire C. Food D. Hill

Drink is to water as eat is to Food
The correct answer is 2.C.

The abstract reasoning test shows a sequence of four simple diagrams arranged in a progressive sequence according to rules of varying obscurity. The student should pick from a selection of five additional diagrams that one which most logically continues the sequence of the previous four.

It would seem that, if any improved performance on these tests resulted from the training given the experimental group, this must have been of a rather general and pervasive nature since none of the training was specifically directed toward questions or problems of the type used in the verbal reasoning or the abstract reasoning tests.

On the other hand, it should be noted that the influence of other teaching sections upon the performance on the test has not been included. As will be discussed in the next section, so far as performance on SMSG geometry is concerned, teaching sections in mathematics completely overweigh the effect of any experimental teaching in the general science classes.

In respect to the material presented in this section, it should be that the results are provocative. They suggest that it is quite possible that a significant improvement in general mental capacity has been produced. This increase may be the equivalent of two or three percentile points and thus (referring to Table 3) corresponding to three, or four months of age increment. However, this conclusion certainly cannot be drawn from the study with any great confidence.

E. APPLICATION OF THE RELATIVE RANK INDEX TO STUDY OF A TEST BASED ON SMSG GEOMETRY

Another exploratory item of the "action"-type in Fig. 1, also based on the "hunch" that there might be a general increase in competence to carry out intellectual processes, involved using the 10th grade SMSG geometry text which the 9th grade students in the experiment would use in their next year's mathematics course. Chapter II of this book contained a review of set theory and also deals with relationship of real numbers on a line.

The "hunch" occurred that possibly the ability to do homework in this book and understand the text material might have been increased by the teaching experiment. Consequently, arrangements were made to obtain a sufficient number of the text books and to give a test by asking the students to work as many problems in Chapter II as they could in a one hour period. The relative ranking indices of the experimental groups discussed in Section D together with the same control groups are shown in Table 12.

TABLE 12. RELATIVE RANKING INDICES RELEVANT TO SMSG GEOMETRY CHAPTER 2 TEST

EA (69) vs C (49) DAT/V+N Fall 1963	-16 *percent*
EA (69) vs C (49) Geometry June 1964	-48
DoA + RA (47) vs C (44) DAT/V+N Fall 1963	-19
DoA + RA (47) vs C (44) Geometry June 1964	-36
DoA + RA (47) vs C (42) Math. Classes	-38

At first glance, it appeared that the experimental training had a distinctly adverse effect. The ranking of students on DAT/V+N in the fall of 1963 was -16 percent compared to the control group. On the geometry, they fell to -48 percent as if the controls had gained 32 percent in respect to the experimental groups.

This result was further analyzed in more detail. Two sections from the experimental groups who seemed to have a combined distribution possibly closer to the control group than did the entire experimental group. Furthermore, some students were eliminated from the control group whose other grade records made it appear that they did not properly belong at this level. This leads to a change in which the experimental group was found to be at a relative rank index of -19 percent compared to the controls. On this basis, the relative rank on the geometry test was -36 percent indicating once more that the controls had grown faster than the experimental group and had gained 17 percent in relative rank index compared to the experimental group.

In the course of this investigation, the idea occurred that the experimental and control groups might be in different mathematics sections. It was found that the mathematics classes themselves could be ranked in a hierarchy of about 9 rankings which were significantly occupied by students from the sample shown in Table 12. There was one group which was clearly the outstanding mathematics class, a second class, two classes tied for third, and then a sequence of six classes of gradually descending mathematical competence.

A new relative rank index was made based on ranking in the mathematics classes. When this was done it was found that the mathematics ranking of the experimental group was at -38 percent compared to the control group. In other words, the control group was definitely of higher mathematics potential and of higher mathematics standing than the experimental group. This indicates that the effect of the experimental training, and insofar as Table 12 shows, may have produced an insignificant improvement of 2 percent over the expected value based upon relative ranking in the mathematics classes.

F. THE LOGICAL STRUCTURE TEST

During the summer teaching session discussed in the Appendix 3 written by Miss Humphrey, an attempt was made to devise a test which would indicate the understanding of logical structures like geometric theorems but in a more general context.

For example, in reasoning pertaining to the checkerboard test discussed in Subsection C, the basic postulate on which subsequent reasoning is based is that a domino covers one square of each shading. From this, the fact that the total number of squares covered by dominos consists equally of shaded and unshaded squares is a theorem. A conservation law is that the difference between number of uncovered squares that are shaded and uncovered squares that are unshaded is conserved and invariant to the number of dominos that are placed on the checkerboard.

A test was designed to see if the students could select which statements in respect to the checkerboard test and other problems were like postulates and which were like theorems or conservation laws. The actual structure of the test is given in Appendix 2.

Once more limitations of time and manpower have prevented a thorough analysis, but a brief survey indicates that on tests like the checkerboard test, and a similar and equally obvious test known as the tennis tournament, about one third of the students could not recognize which statements were like the postualtes.

In keeping with the comments made by Mr. Millar in Appendix 7, the Vice-Superintendent of Cuberley High School where the tests with 10th grade students were carried out, this test gave the feeling that an important aspect of the subject matter is that it involves logical thinking and may be more closely related to mathematics than to science. In a future study, it is thought that introduction to the logical structure in the games and mathematical problems and the law of levers and other manners related to conservation of energy should be brought out much more overtly and students taught to look for such logical structures.

G. APPARENT AGE SHIFT

In this section we summarize the effect of percentile shift on experimental group EA of Table 1 and 7. As discussed in connection with Table 3, we see that 10 percentile points at 85 going to 95 for 9th grade students is roughly equivalent to one year of increased experience while 14 percentile points from 85 to 99 is approximately equivalent to 2 years.

The results discussed in the previous sections are summarized in Table 13. We see here that taking the median of lane A as the 87th percentile for the differential aptitude test based on verbal plus numerical scores, the shifts vary from 10 to 12 percent on these scores. On the other hand, if we use an abstract reasoning score to compare the checkerboard test with the CO control group of 11th and 12th grade science students, in Ohio, we see that on abstract reasoning this would amount to an increase from 87 percent to 95 percent.

TABLE 13. APPARENT UPWARD SHIFT ASSOCIATED WITH EXPERIMENTAL TEACHING

CHECKERBOARD 2	Percent
EA vs JF - Match of median test scores implies shift median DAT/V+N:	87 to 97
EA vs JA - Approximately 32 percent of EA pass median of JA:	87 to 99
EA vs CO - Upward shift of about 2.5 *years*. *Equivalent* on DAT/AR to	87 to 95
VR AND AR RELATIVE RANKING INDICES	
EA vs all C in AR + 13 percent:	87 to 89
EA vs C in VR + 2 to 10 percent:	87 to 87-89
EA vs C in AR + 13 percent:	87 to 89

The percentages shown in the upper part of the table based on checkerboard 2 are larger than those shown in a lower part of the table based on verbal reasoning and abstract reasoning. This is certainly consistent with the experimental background since the checkerboard test is more closely related to the material taught than the verbal reasoning and abstract reasoning which should measure more general attributes. The verbal reasoning and abstract reasoning suggest an effect of the order of 2 percentile points in rank at the 87th percentile, or about ⅕ of a year of increasing experience.

In making these comments, it should be realized that an attempt has been made largely to show the proportion of the numbers in respect to other scales than to suggest that a real causal effect can be claimed at the present time.

An attempt to make a similar appraisal for the lane B group has not been made due to lack of manpower and time.

H. NOTE ADDED IN FINAL DRAFT

The abstract reasoning data has also been analysed, at the suggestion of Prof. L. E. Moses, by a covariance test utilizing raw scores before and after training. The results are in keeping with those quoted in Table 11. In the analysis the students were divided into three groups corresponding to their mathematics classes. The results are as follows:

TABLE 14. RESULTS OF ANALYSIS OF COVARIANCE FOR DAT AR.

Math Class	No. Students Exp.	Control	Extra Problems Solved by Exp.	F-Values Parallel Regression	Effect
GB	6	10	-1	0.00	< 1
GC	19	10	+1	0.00	2.5
M1	12	9	+4	<1.0	9.6
Total	37	29			

This analysis shows that no significant effect is observed for the students in the two more advanced mathematics section GB and GC. A significant advance of 4 problems (this is more than one years normal advance) was found for the students in M1 at a significant level of less than 1 percent.

These results are in keeping with the results of Table 11 since one year for about one third of the group corresponds to the three or four months for the group as a whole as concluded in the last paragrah of Subsection IIID.

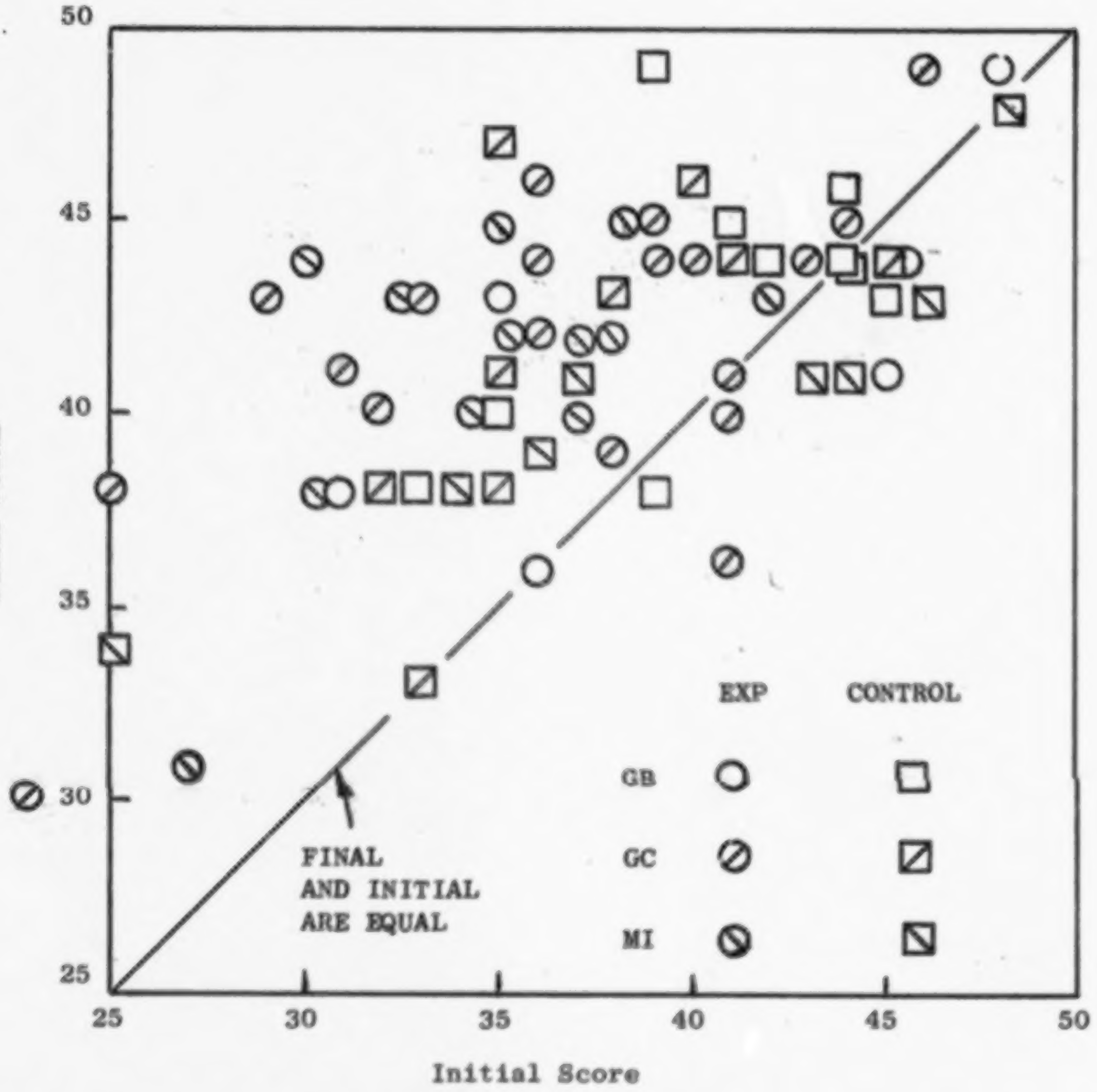


FIG. 2. DATA FOR TABLE 14. This shows the final scores on the DAT Abstract Reasoning test plotted vs initial scores for the experimental group (circles) and the control group (squares). Improvement, represented by displacements above the "equal" line, are seen by inspection to be greater for the experimental group than the control group.

IV. CONCLUSIONS

Two pieces of evidence suggest that the educational experience affected cognitive development. First, on the Checkerboard Test, experimental subjects equalled the performance of control subjects who were presumably more advanced in their cognitive development. They equalled the performance of a group, averaging about two and half years older, and who were also selected for their achievement in science. They also, on the same test, equalled the performance of an Advanced Placement Class in science and surpassed the performance of another group which was twelve percentile points on the average above them on aptitude measures. These latter two groups were the same age and in the same school system.

Second, the upper fifty percent of the experimental students advanced significantly and substantially on the Differential Aptitude Abstract Reasoning Test.

At the beginning of the study it was not known what measures on the Differential Aptitude Test were most closely related to the kinds of changes the experimental program attempted to produce. Consequently, the Verbal Reasoning test was also used on the general assumption that it measured a reasoning factor. No significant changes were found on this measure for any of the experimental groups.

No results were obtained on a measure of ability to learn new material and work problems in mathematics. Given a chapter and problems in one of the SMSG mathematics texts, the experimental subjects were not significantly superior to a group comparable in ability and mathematics experience.

In general, changes were not obtained on some tests thought to measure general cognitive functions. However, on those measures that seemed to tap the effects of training, changes equivalent to a half year to two and half years in cognitive development were obtained. It is important to note that these results are suggestive rather than conclusive. Failure to obtain more conclusive evidence resulted in part from the lack of sensitive measurement devices and in part from difficulty in establishing a research design that permitted an unambiguous conclusion.

Because of the difficulties in establishing a tight research design, other factors may be adduced to explain those results obtained. First, the "Hawthorne effect" may be operating. Experimental subjects clearly knew that they were in an experiment; they knew new materials had been introduced and that new ideas were being tried out on them. If there is a Hawthorne effect operating, it ought to appear on all the tests, such as the Verbal Reasoning, and not just on the Abstract Reasoning test. Since the experimental subjects appear superior on only some of the tests, the likelihood of the Hawthorne effect producing these few differences seems small.

The educational experience was conducted in part by the senior investigator, an eminent scientist. His lecturing may have been the factor produc the differences. An independent evaluation of student reaction showed that this lecturing did not receive the same positive evaluation that other aspects of the experimental program were accorded.

The special qualities of the other teachers could not account for the differences since the teachers themselves were learning the material and ideas of the experimental program. If anything, the difficulties the teachers experienced introducing new material would have inhibited the learning of the experimental subjects.

In view of the limitations experienced, the results are probably better than could be reasonably hoped for.

A number of observations were made that are not properly conclusions of the study but are worthy of discussion. First of all it is obvious that teaching the logical structure of science is an activity for which teachers and students are not adequately prepared. It is not routinely done; neither is it expected. Yet, conveying some notion of this structure probably is necessary for the development of an adequate understanding of science and for the development of original and creative thinking in science. In part, some of the lack of results may be due to the unfamiliarity of teachers and students with this approach to critical thinking.

Second, it is obvious that people with training in science and education have not grasped the fundamental significance and utility of basic laws such as that of the conservation of energy. As a consequence a wide variety of scientific concepts are not integrated and explained in terms of such laws. As a matter of fact, the experience of doing this seemed to be quite novel for both teachers and students. The practical consequence is that much science teaching is probably fragmentary. Again, some of the difficulties of obtaining results in this study may be related to the difficulty and novelty of comprehending this way of approaching science.

Interestingly, a number of the experimental subjects were able to discover for themselves one facet of the law of conservation of energy while working on a physics problem. This they were able to do by transfering the approach they had learned by playing the logical games. Consequently, these "successes" are evidence that the effects predicted for the experimental program can occur, and that an enlargement of the procedures used will produce more significant results.

A number of educational procedures were used that appear promising in a program of this kind. One of these, the logical games, is particularly interesting. Students found the games both challenging and instructive. They are an easy device for making significant logical points. It appears that students can initially grasp the points more easily in the context of the game than they could if presented in the context of science material.

Another procedure used was to have students teach each other. This procedure was liked by the students. They felt that they learned more by teaching. This procedure makes a tutorial method possible, and one is required to make sure these concepts.

A number of testing devices were also invented which are necessary to developing the kinds of tests required. One of these was the use of sequential hints to see how much prompting was required before a student solved a complex problem.

In summary the study has shown that given appropriate conditions, refinement of procedures, and sensitive testing devices a significant demonstration of the effects of the program almost certainly could be achieved. Further, the experience of conducting the study has revealed weaknesses in the educational program which can be remedied only by an educational experience similar to the one used in this study.

APPENDIX 1

EXPERIMENTS

(A complete collection of the relevant
material for the Wilbur experiment
has been recorded on 35 mm microfilm)

Dr. Shockley

February 7, 1964
Introduction

EXPERIMENTS ON HANGING A RIGID BODY

When an airplane loses a propell~~o~~er blade and the resulting vibration tears the engine from the wing, it is because the balance has been destroyed. Many automobiles and especially sports cars have little weights attached to their wheel rims; these weights balance the wheels and reduce vibrations. When children sit on a see-saw or teeter-totter just right, it balances.

All of these different balancing effects depend on one principle. Every large body has a center of weight. There are certain principles that apply to this center of weight.

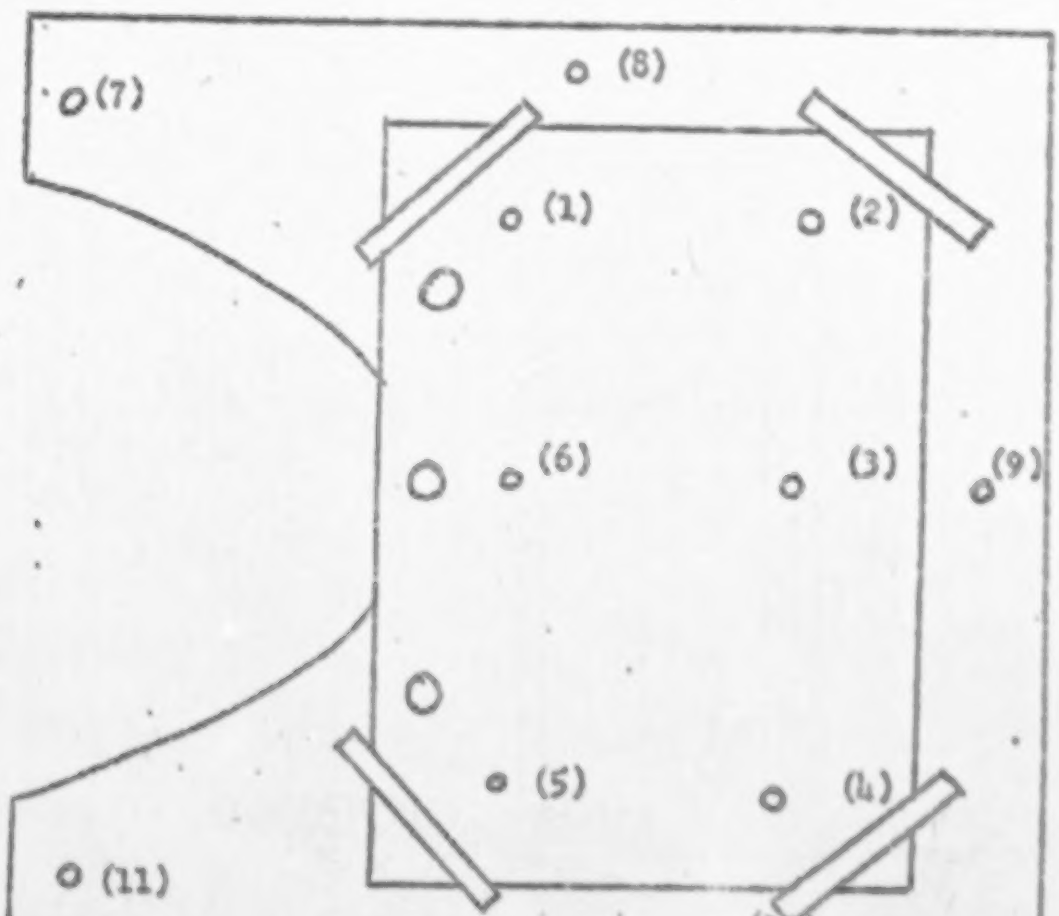
When you understand these principles you will find that there is a correspondence between them and the principles in the three person match game of take one or four. In the match game you concentrate your attention on the remainder after dividing by three; by studying this remainder, you understand the principles of the game. By concentrating your attention on a particular point of the plywood board in this experiment, you can understand the principles that determine how it acts under many circumstances.

EXPERIMENTAL APPARATUS

The apparatus involves chiefly three things: (1) an irregularly shaped rigid board of plywood with a set of holes bored around its edges to act as pivot points from which it can hang; (2) a horizontal rotation axis made by driving a nail into a narrow piece of board which can be held out from the edge of the top of a table or desk by weighing it down with books; and (3) a plumb line consisting of some fishline and a plumb-bob consisting of a fishing sinker.

Your board will have holes numbered 1 to 6 which will be covered by the graph paper and also hole F (stands for Fulcrum). Holes numbered 7,8,9,10, and 11 do not lie under the graph paper.

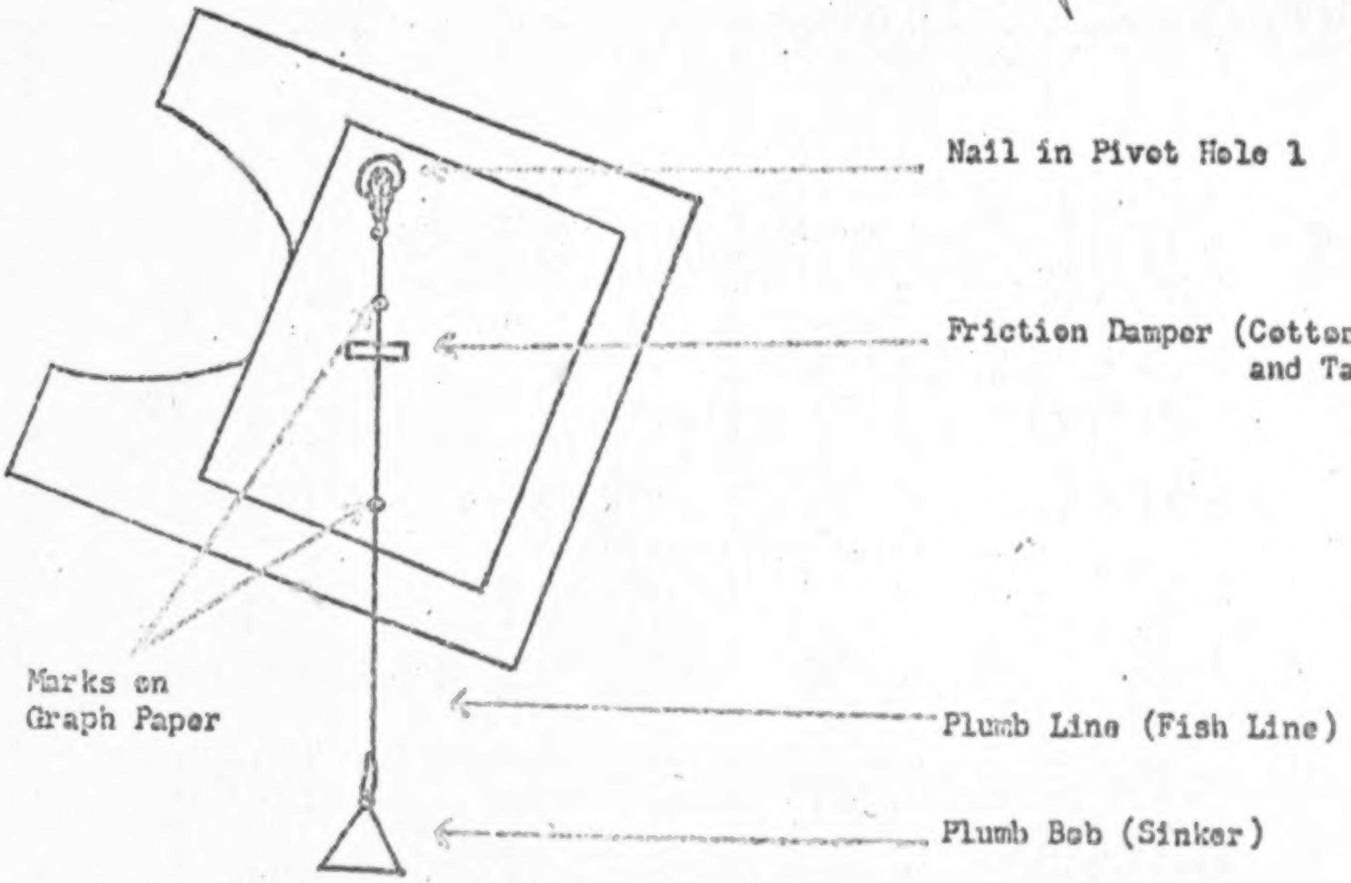
Mount the graph paper as shown below. Hold the board up to the light, then push a nail or pencil through the paper into holes 1 to 6 inclusive. DO NOT PERFORATE PAPER OVER HOLE F. Perforate paper from paper into board, not the other way. This keeps the edges of the holes neat.



Pivot Line 1:

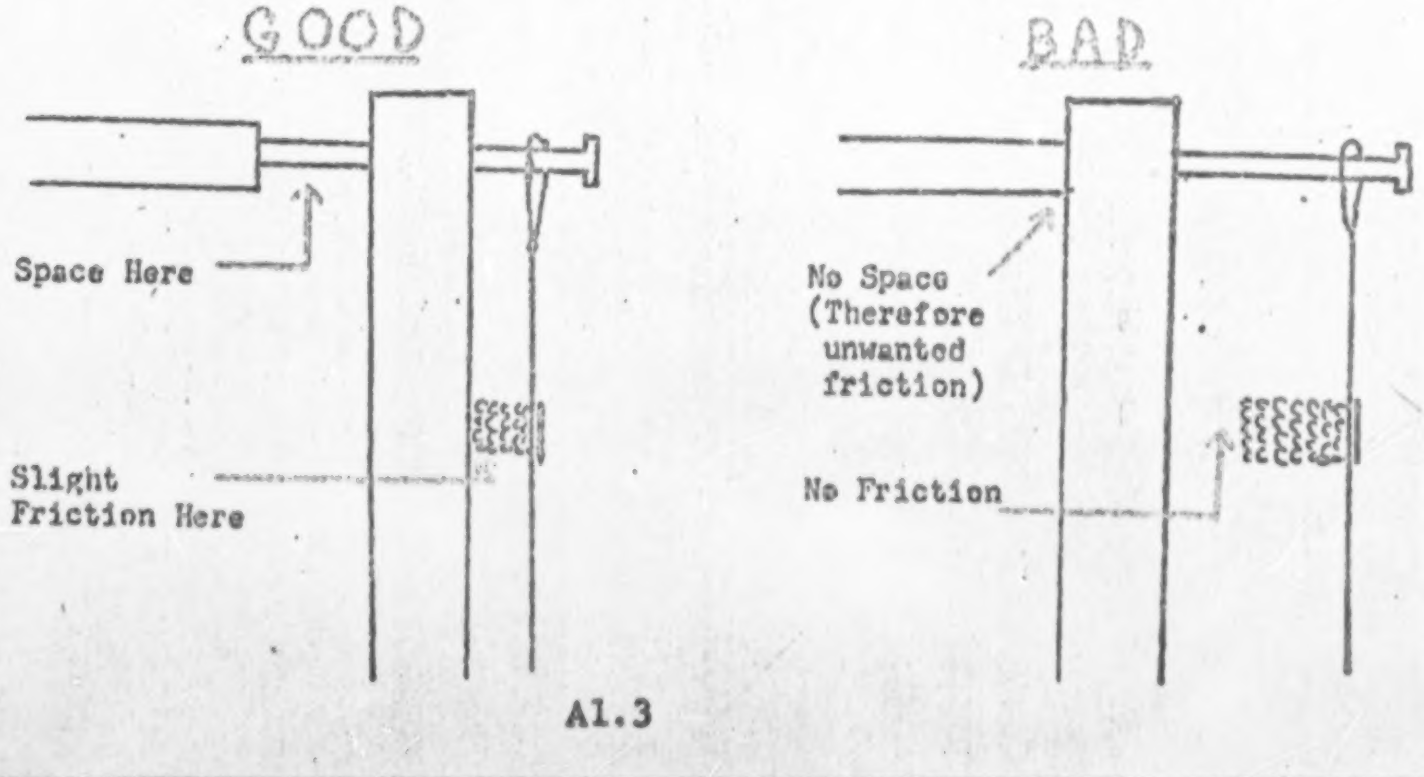
Hang the board from the axis nail through pivot hole 1 with the graph paper out. Hang the plumb-bob line in front of the graph paper. Adjust the cotton wool damper to stop the swinging gradually. It should look like this:

tie frictionless knots

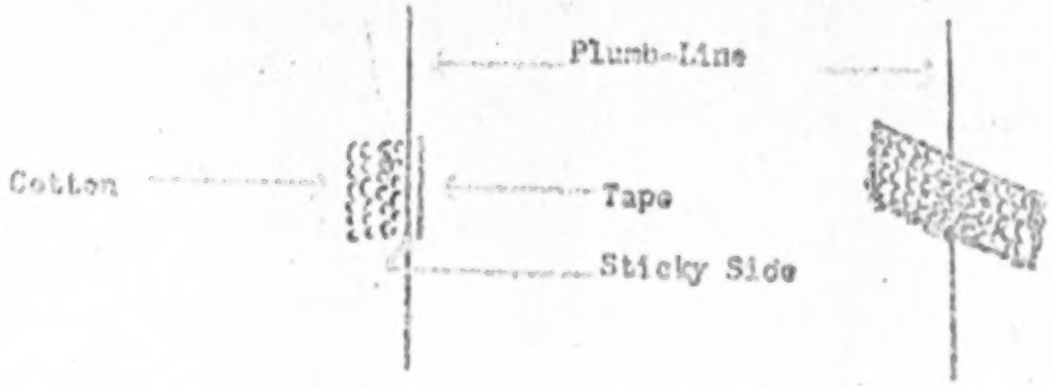

Nail in Pivot Hole 1
Friction Damper (Cotton Wool and Tape)
Marks on Graph Paper
Plumb Line (Fish Line)
Plumb Bob (Sinker)

keep free

Viewed from the side it should look like this:


GOOD
Space Here
Slight Friction Here
BAD
No Space (Therefore unwanted friction)
No Friction

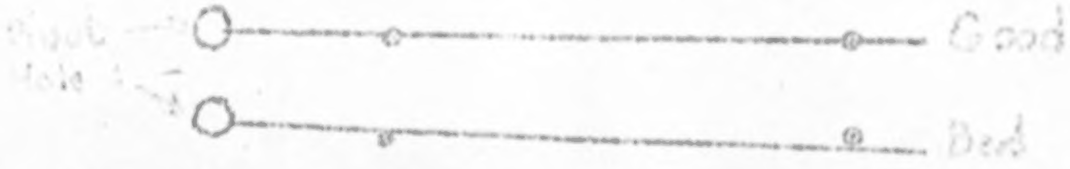
If you don't have a damper, make one by getting about a one inch length of scotch or masking tape and pressing it against the plumb-line from one side and the cotton wool from the other.



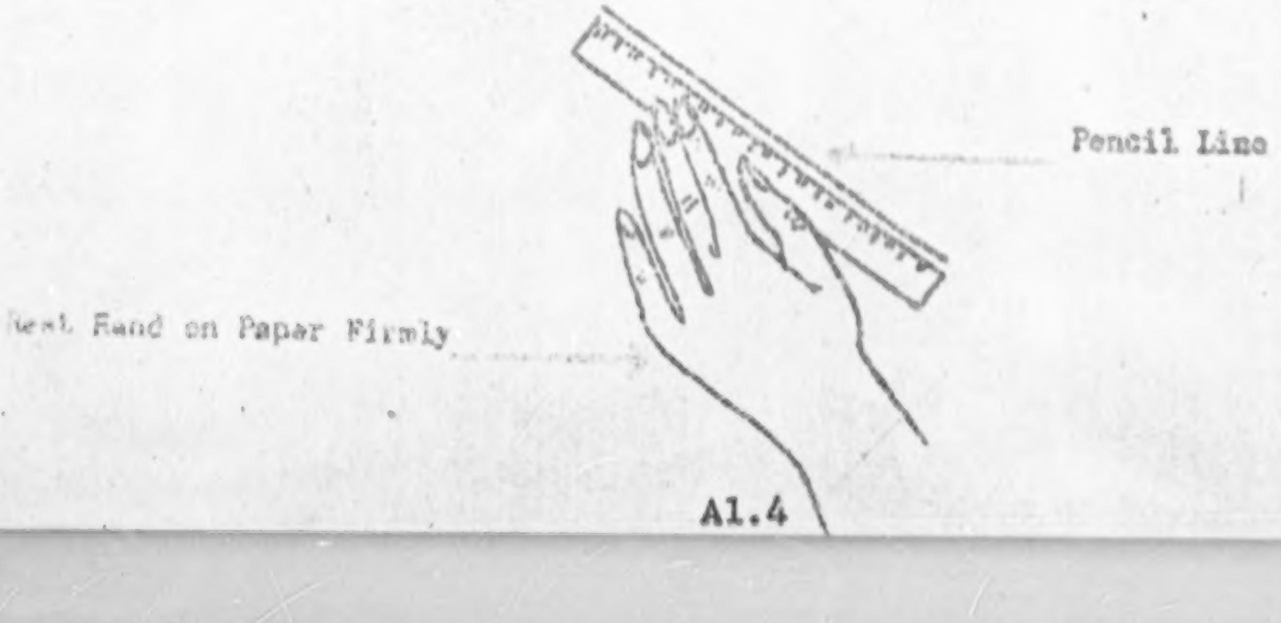
Before you determine the pivot-line from pivot hole 1, stop by hand any violent swinging of the plumb bob. Then, after half a minute or so, the damper will bring the plumb-line to a stop against the board. Look perpendicularly towards the graph paper and fix your attention on a point on the paper (select one that is easy to keep track of) directly behind the plumb-line. Push the plumb-line aside and mark this point with your pencil. Now bring the plumb-line to rest again and check the accuracy of your mark. Repeat if your error is greater than 0.5 mm. (BE SURE YOUR PENCIL IS SHARP)

Make another pencil mark higher up on the graph paper as shown on the figure of the hanging board.

Next, take the board off the pivot nail. Hang the plumb-line back on the nail to keep it ready for use. Take the board to a table and CAREFULLY and ACCURATELY draw pivot line 1 through your two marks. Continue your line right up to pivot hole 1. It should look like this:



To avoid a BAD line, do not let the ruler slip; hold ruler firmly in a fixed position on paper like the diagram below. Test to see how far away from the edge of the ruler your pencil line will come. Adjust so that your line goes right through both marks with an error of no more than 0.5 mm.


Pencil Line
Rest Hand on Paper Firmly

Pivot Line 2:

Repeat same procedure for pivot line 2, using pivot hole 2.

Hunch:

Predict where pivot line 3 will go. Draw it in as a dashed line and label it "Predicted pivot-line 3".

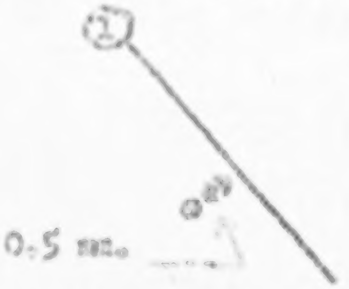
Pivot Line 3:

Repeat same procedure and accurately draw pivot line 3.

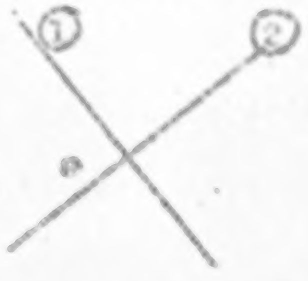
ANALYSIS, HUNCH, and the HYPOTHETICAL MAGIC PIVOT-LINE POINT:

Do you have a HUNCH that all the pivot-lines should go through a common point, which could be called the "magic pivot-line point", and which mysteriously makes the plumb line pass through it no matter how the board is hung? Of course, this would ideally require perfect roller-bearing pivots and not just a nail through a hole in a board. Furthermore, your lines would have to be drawn perfectly and you never could make them fine enough so that the intersection is really an exact point. But suppose there is a magic pivot-line point for an ideal system; what could you expect of your imperfect experiment?

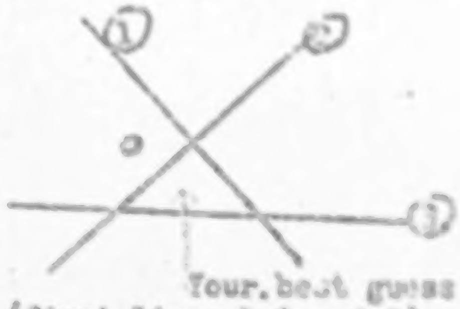
Suppose all the imperfections make your lines have errors of 0.5 mm., then what would you expect if there really were a common pivot-line point? You would expect that if you had bad luck and the errors added up, your lines might miss by as much as about 1 mm. Suppose the lines miss like this:



(Pivot-line 1)



(Pivot-lines 1 and 2)



(Pivot-lines 1,2,and 3)

Then you might guess wrong by about 1 mm but if you try more lines it should average out.

VERIFICATION OF HYPOTHESIS; EXPERIMENTS ON OTHER PIVOT HOLES

Check your hunch next. Hang the board from the other pivot holes and observe that the magic pivot-line point always hangs directly below the point of support at the nail.

PRINCIPLE FOR THE CENTER OF GRAVITY:

Your experiments show that there is a magic pivot-line point on the wooden board which always tries to hang directly below any of the pivot holes.

This point is understood in terms of the concept of the pull of gravity upon all of the separate molecules making up the board. The effect of this pull of gravity is as if all of the weight of the board were concentrated right at the magic pivot-line point.

If all of the weight of the board could be drawn together and put at this point, then hanging the board from any one of the pivot holes would be just like hanging a plumb bob from a line or as it is called, a pendulum. The point where the weight is concentrated would then hang directly below the pivot line point.

The fact that the pull of gravity on the board makes it act as if all of its weight were centered in this one pivot line point is the reason that this point has the name THE CENTER OF GRAVITY.

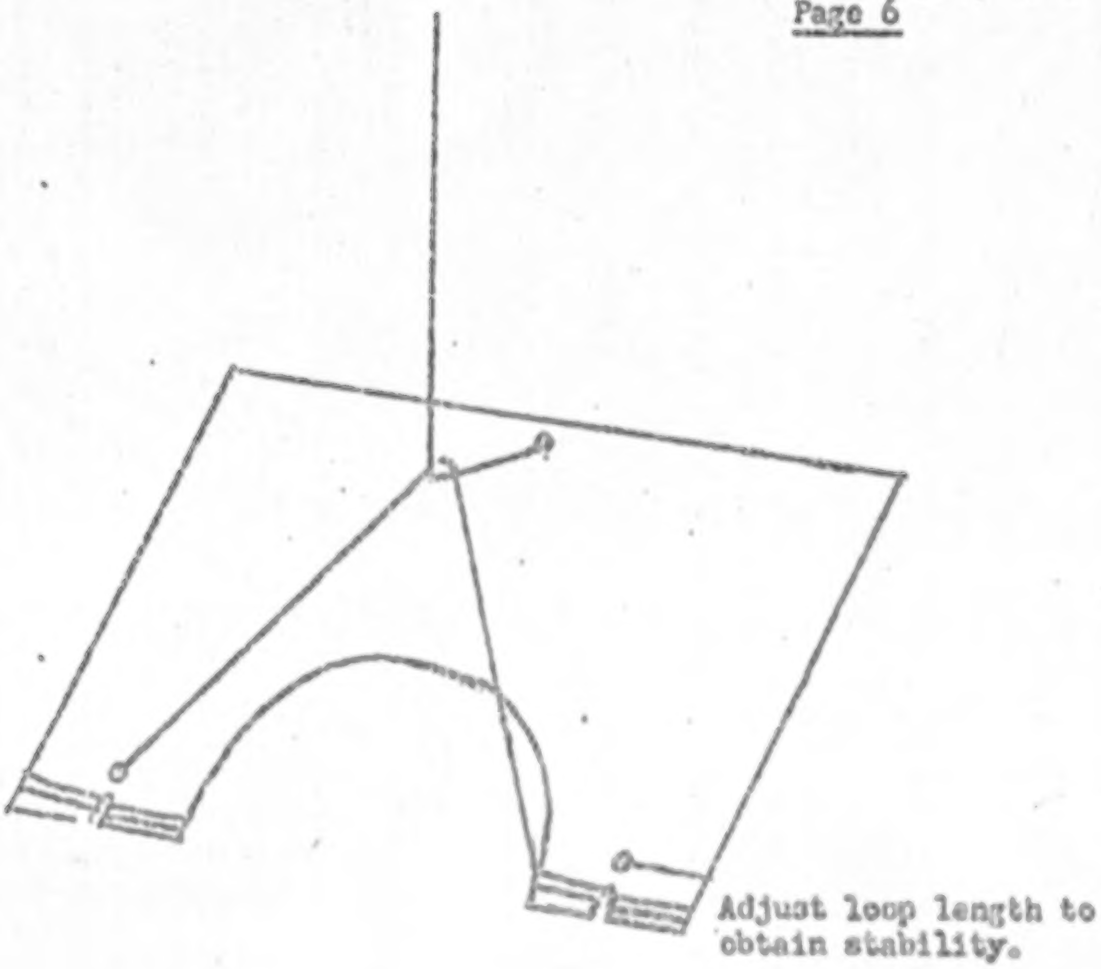
PIVOT HOLE F:

So far you have hung the board only from holes which are near the edge and very far from the center of gravity. You will find that there is another hole in the board under your paper, which is marked F, standing for Fulcrum, on the back of the board. Punch the paper over this hole and hang the board from Pivot Hole F. You will find that the center of gravity point tries to come to where it hangs directly under the pivot.

HORIZONTAL POSITION: (This section to be written more fully later.)

If the center of gravity always hangs below the point of support, it should not matter how it is hung.

Hang the board on the "pyramid" suspension. (See figure. Teacher will show you how, if necessary.) Sight down the string and see that center of gravity is below support. Readjust pyramid support and check again.


Adjust loop length to obtain stability.

RESULT:

How would you state the principle you have found about the center of gravity? Put it in Standard Form for a Principle.

EXPERIMENTS ON HANGING A RIGID BODY THAT SUPPORTS MOVABLE WEIGHTS

After this experiment, turn in the graph paper and the page of answers.

STEP 1: Mount a fresh piece of paper on the plywood board. Punch through pivot holes 1,2,3,4,5 and 6 and number them. Write your name, teacher and date in the upper right hand corner.

STEP 2: In the usual way using the plumb-bob, locate and draw pivot lines 1, 2 and 6. Mark each of these lines "S2" so as to separate them from Step 3. Put a circle around their intersection; label it "COS2".

Write your name, teacher and date on another piece of paper and answer Question 1.

QUESTION 1: How accurately do your pivot lines cross? About how accurately have you found the center of gravity of the board?

STEP 3: Hang a 3 oz. weight by rubber bands from the two "horns" of the plywood. Hang the board from hole 6; the weight should come close to the edge of the graph paper but not touch it. You can adjust this by changing the rubber bands or winding them around the horns. Locate a new pivot line 6 and also new pivot lines 1 and 2. Label these "S3" so as to distinguish them from the "S2" pivot lines. Put three small circles around the three points where the S2 lines cross. Label them S2 also.

QUESTION 2: Is there a magic pivot line point that all the pivot lines go through?

QUESTION 3: Does the board have a real center of gravity?

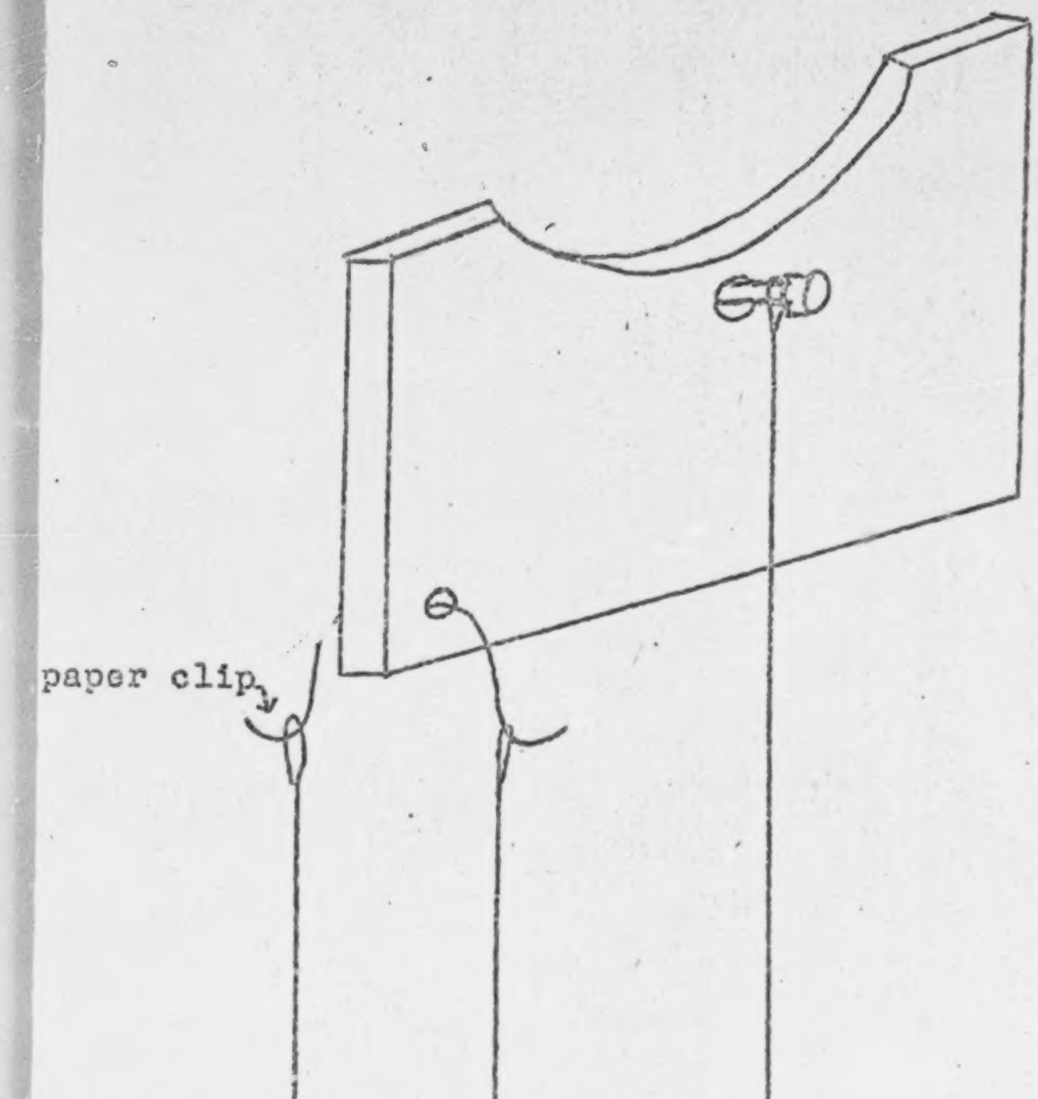
QUESTION 4: If the rubber bands were replaced by metal wires that held the 3 oz. weight in a fixed position in respect to the board. what would happen? Would there be a magic pivot line point which would be the center of gravity?

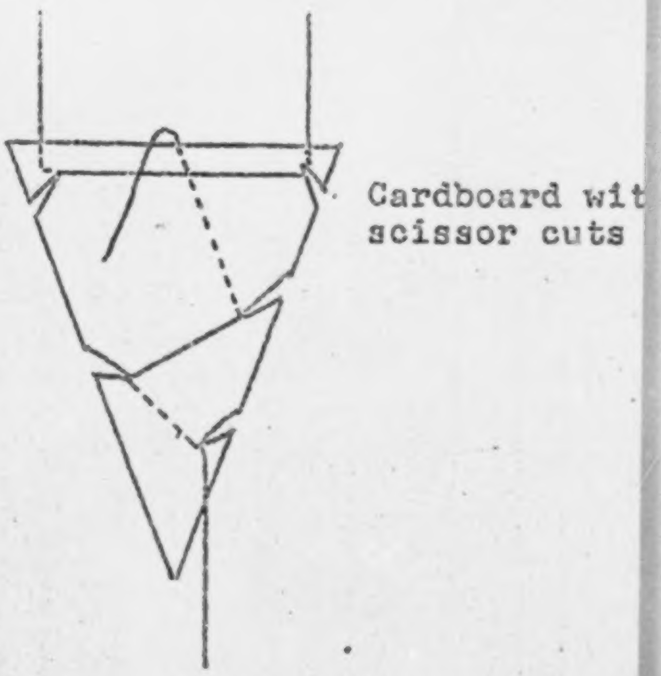
STEP 4: Hang the board on Pivot hole 1 and then hang a 3 oz. weight on a bent paper-clip through pivot hole 4. See the diagram. Construct a new pivot line 1 for Step 4. Mark it 'S4'. Change the length of the string to the weight so that it hangs at a different distance below the board.

QUESTION 5: Does the position of pivot line 1 depend on the length of the string?

Step 4(cont.) Construct new pivot lines 2 and 6 with the 3 oz. weight hanging from pivot hole 4. Mark them "S4".

QUESTION 6: Is there a magic pivot ~~line~~ through which all the pivot lines want to go? Call this point "COG4"?


paper clip
oard with
punched


Cardboard wit
scissor cuts

QUESTION 7: How can you think of the effect the 3 oz. weight hanging from pivot hole four has on the board? Imagine that the 3 oz. weight were compressed into an imaginary nail and stuck into pivot hole four. Would this have the same effect as the weight hanging from the harness?

QUESTION 8: If there were an imaginary very heavy nail in pivot hole 4 that weighed just as much as the 3 oz. weight, would you expect the center of gravity of the board-nail combination to lie on a straight line running between CGS2 (the center of gravity of Step 2) and pivot hole 4?

QUESTION 9: What does your graph paper data show about the three points: CGS2 (the center of gravity of Step 2), CGS4 (the magic pivot-line point of Step 4) and pivot hole 4? Draw a dashed line between "CGS2" and pivot hole 4.

QUESTION 10: Is CGS4 really a center of gravity for the board and the hanging weight, or does it just act like a center of gravity?

QUESTION 11: If a pivot hole were drilled right through "CGS4" do you think the board with the hanging weight would balance there?

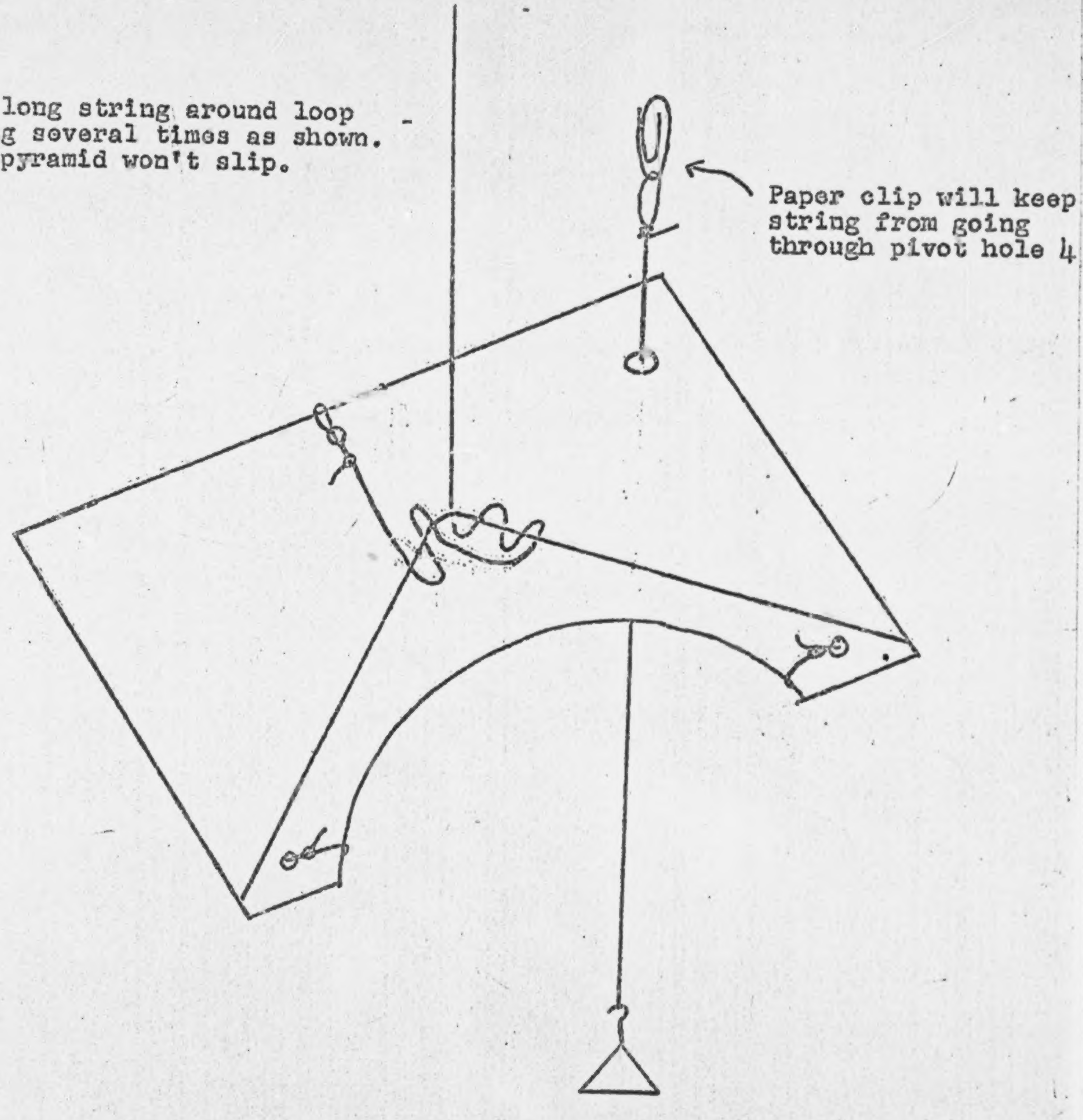
QUESTION 12: Imagine that the board were hung from pivot hole 1 with the 3 oz. weight hanging from hole 4; then imagine the hanging weight were fixed rigidly in position in respect to the board by turning the strings into imaginary steel rods that could not turn in pivot hole 4. Would the board balance in all positions on a pivot hole bored through CGS4?

STEP 5

Hang the board from the pyramid suspension and hang the 3 oz. weight from hole 4 with a single paper clip. You will have trouble with the pyramid threads slipping unless you wind them around like the diagram (see next page). If you wind them like the diagram you can slide them where you want, but they will not slip.

QUESTION 13: Do any of the "CG" points line up with the vertical string?

QUESTION 14: What principles can you state from these experiments?



Wrap long string around loop string several times as shown. Then pyramid won't slip.

Paper clip will keep string from going through pivot hole 4

26 February 1964

HOMEWORK EXPERIMENT ON CENTER OF GRAVITY

This experiment is intended to show how the position of the center of gravity can be calculated for simple cases. In particular, it shows that the center of gravity of two bodies of equal weight always lies at the midpoint between the two centers of gravity of the two individual bodies.

The equipment required for the experiment is a piece of cardboard approximately 8½ x 11", a metric system ruler 12" long, a pair of scissors, and a pencil.

Write your results on a piece of paper under the headings Section 1, Section 2, etc., as described below. Turn this in with your report. Also turn in the larger piece of cardboard discussed below.

First mark the piece of cardboard into sixteen rectangles as shown in Fig. 1. Label the rectangles with letters and mark the centers of the other rectangles with numbers as shown. Also number the places where the solid lines of the numbered rectangles hit the edge of the big numbered rectangle. This is for use in description later. Mark the centers of the numbered rectangles as shown and label the center of the big numbered rectangle as CG. This is the center of gravity of this piece of paper.

FIG. 1. Cardboard marked for use.
Cut along lines marked —o—o—o—

© 1964 William Shockley

After you have completed the marking, cut out all of the lettered rectangles, leaving the big numbered rectangle as one piece. Be careful to cut accurately along your lines, because if the little rectangles are of different size they will have slightly different weights and this will introduce errors into the experiment.

Step 1

First check your center of gravity by balancing the large numbered rectangle on the edge of a ruler. To do this you may hold onto one corner of the rectangle and slide it over the ruler until it just tips one way and then just tips the other way. Then you may hold it halfway between these positions with one finger, as represented in Fig. 2 (which shows how it looks in Step 3), and mark the places where it lies on top of the ruler with a pencil as shown. Draw two diagonal lines on the rectangle by this means and see how close they come to the center of gravity.

Report Section 1. Write a brief description of how you have done this and label the point or the cardboard which you found in this way as Step 1.

FIG. 2

Step 2 and Report Section 2

Balance the board on lines (3 5) to (4 6) and see how close the balance is to this line. Repeat this on lines (1 2) and (7 8) and describe your results on your answer sheet as Report Section 2.

Step 3

Stack the eight little rectangles on the center of square 2. They should weigh just as much as the big rectangle and their center of gravity should be in the middle of rectangle 2. Use your centimeter ruler and divide the line between CG and the center of 2 in half. This point should lie (3/4)H above the center of gravity on Fig. 1 and (1/4)W to the right of the center of gravity. Draw a horizontal line through this new point and a vertical line through this point on the big piece of cardboard. Check the balance and see if this point is really the center of gravity.

Report Section 3. Describe the results obtained in Step 3.

Step 4

In this step you will verify that if two bodies B_1 and B_2 have a center of gravity and they are combined with a third body, then the center of gravity of the three bodies can be found, as follows:

First you can find the center of gravity of the two bodies; then you can consider this to be a single point and find the center of gravity between it and the other body. To test this in one case, put four of the lettered rectangles on the center of rectangle 2 and the other four on the center of rectangle 8. These two stacks have their center of gravity on line (3 5) (4 6) at a distance of (1/2)W to the right of the CG. This pair weighs as much as the big rectangle. The center of gravity of this pair and the large rectangle should then lie at (1/4)W to the right of CG. Check whether or not this is true by balancing along the vertical line used in Step 3 at (1/4)W to the right of CG and the horizontal line through CG.

To test this principle again, make the following arrangement: four rectangles on 3, two rectangles on 4, and two rectangles on 8. The two on 4 and the two on 8 have a center of gravity at the middle of 6 and act like four rectangles at 6. This center of gravity at 6 and the four at 3 combine to make a center of gravity midway between 3 and 6. But this is the same as the center of gravity for the large rectangle. Test that this conclusion is right.

Report Section 4. Discuss what you found here.

Step 5

Arrange the eight lettered rectangles as follows: one rectangle on 1, two rectangles on 3, one rectangle on 5, and four rectangles on 6.

Report Section 5. Where does the center of gravity of this arrangement of little rectangles and the big rectangle come? How can you explain this in terms of the combinations of centers of gravity?

Step 6 and Report Section 6

Invent a new experiment of your own using the eight lettered rectangles; arrange these in some way on the numbered rectangle; predict where the center of gravity should be and test it experimentally. Describe your arrangement, how you made your prediction, and what your experimental result was. After you have done this, attach the little rectangles to the big rectangle with scotch tape and turn this in as part of your report.

Step 7 and Report Section 7

This "brainbuster" problem is intended to give you a hint about the underlying unity of mathematics and science. Suppose you had four equal weights A B C and D fastened to a weightless rigid board. (If you prefer, you can think of A being the center of gravity of the board, which weighs just as much as each of the other weights.) Represent this by putting four dots on a piece of paper at the corners of an irregular four-sided figure like that shown in Fig. 3 (don't use a square or a rectangle). Now use your ruler to find the center of gravity of A and B and mark this point (AB); then the center of gravity of C and D and mark it (CD); then find the center of gravity of the combination (AB) and (CD) and mark this (AB,CD).

Now find the center of gravity in another way: First find the center of gravity of A and D and then B and C. Then find the center of gravity of (AD)and (BC) ; call this (AD,BC).

What do you observe about the two constructions you have found to get the center of gravity of the four weights?

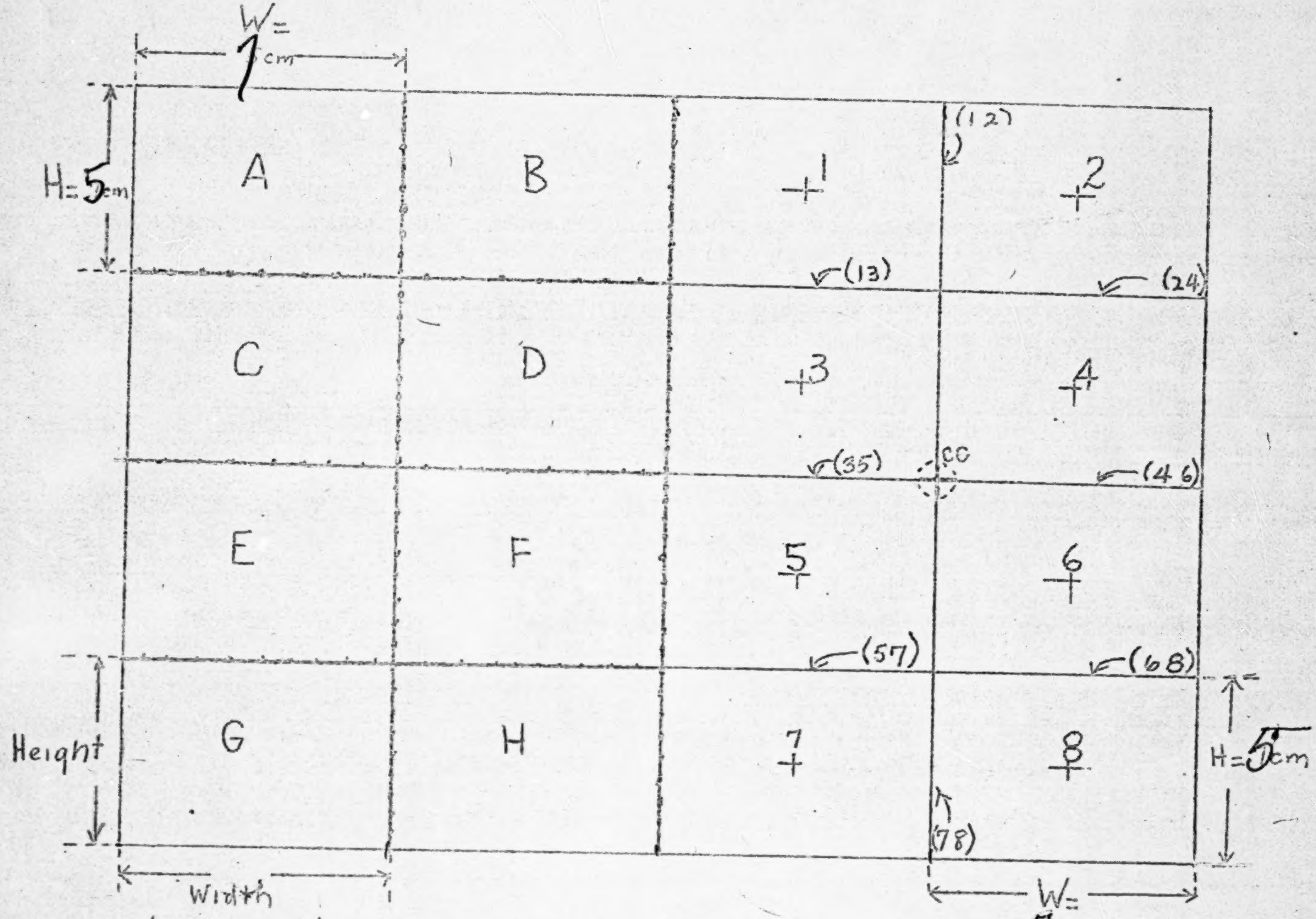
Consider a possible geometrical theorem:

> "Given a quadrilateral, find the midpoint of each side. Draw lines from each midpoint to the midpoint of the opposite side. Then these two lines bisect each other."

Do you believe this theorem is correct?. Why?

What relationship do you see between this theorem and the existence of a center of gravity?

A116


W= 7cm
H= 5cm
A
B
1
2
(12)
(13)
(24)
C
D
3
4
(35)
CC
(46)
E
F
5
6
(57)
(68)
G
H
7
8
(78)
Height
Width
H= 5cm
W= 7cm

Overall dimensions of
Cardbeard - 20cm × 28cm.

Fig. 1
Cut along lines marked

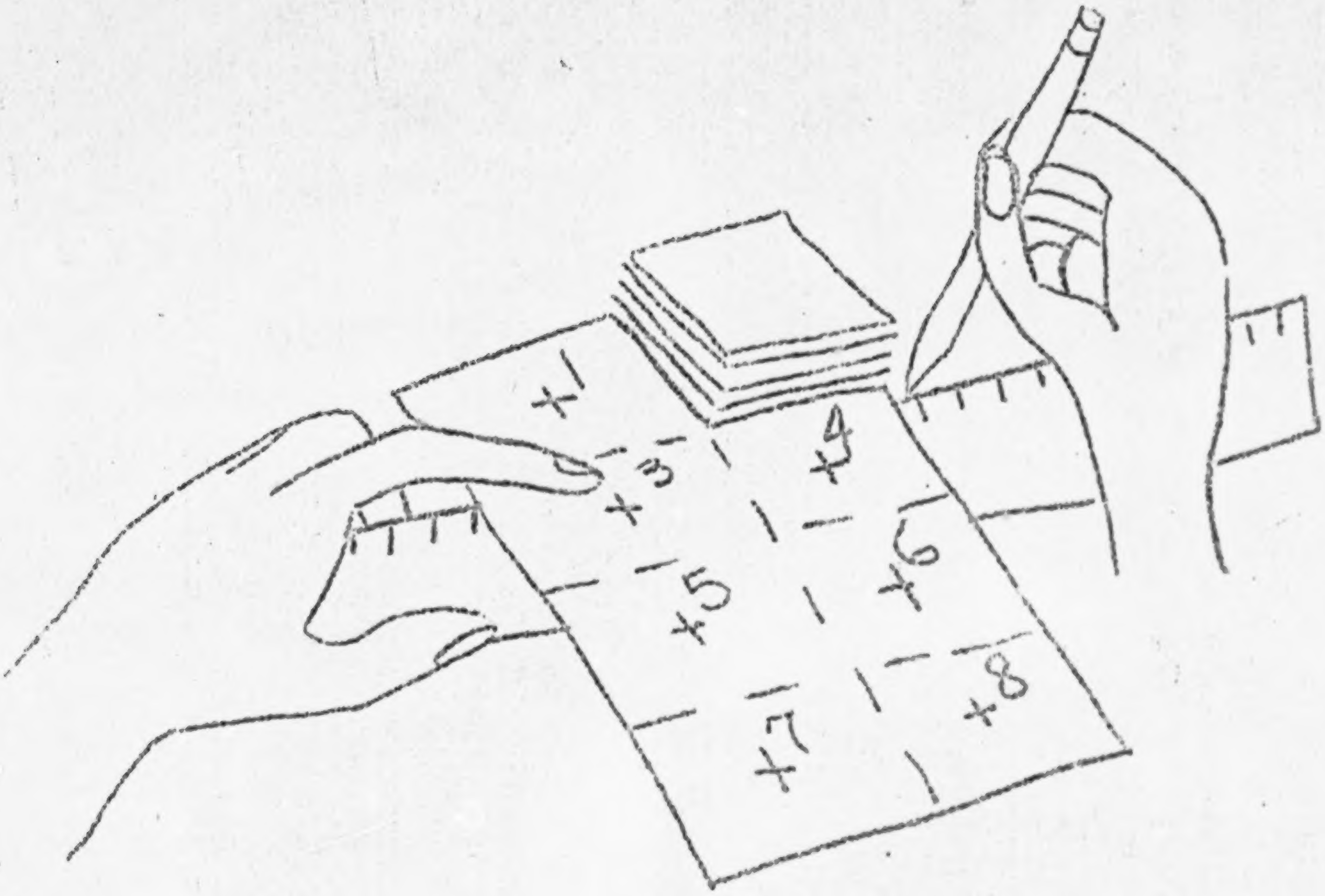
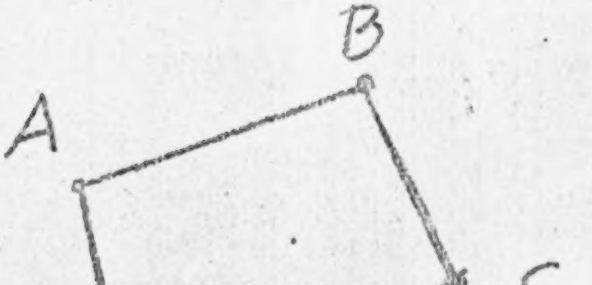
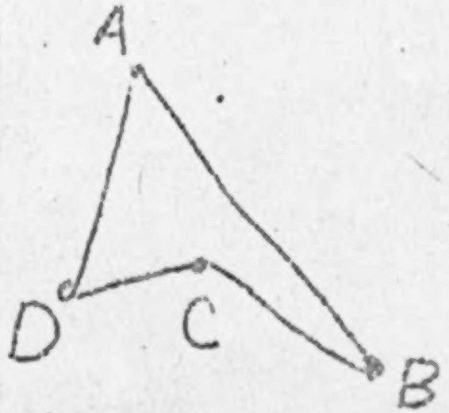


FIG. 2





Dr. William Shockley

February 26, 1964
Introduction

In this experiment you will have an opportunity to discover for yourself the law of moments which is the principle that governs levers, balances and such mechanical machines as the crankshaft of your car. The moment is what you have probably found out was important in making a see saw balance. But what is it exactly? Understanding what the moment is and focusing your attention on it is like discovering the remainder in the take one or four game.

©William Shockley 1964

EXPERIMENT ON LAW OF MOMENTS

Your written report of this experiment on the law of moments will consist of six parts. Each part should be labeled according to its designated number 1. Example: Report - Section 1.

PROCEDURE:

Step 1 Make the centimeter ruler from graph paper as shown in Figure 1 and mount it on the plumb line with scotch tape, as shown in Figure 2.

Each section of the graph paper is to be 25 cm. long. Trim the graph paper evenly at the ends as in Figure 1.

There is an overlap of 4 cm. where the 2 sections are joined in the middle. Each part extends 2 cm. beyond the zero position in the overlap part and should be numbered out to 22 cm., with 1 cm. left over. The ruler should be nearly perfectly balanced when mounted on the plumb line so there is equal weight on either side of the zero position. It is ~~not~~ most important to have the plumb line accurately on the zero mark.

Step 2 Balance the board by hanging it from the fulcrum hole and using 2 washers, as you have done in earlier experiments. The board is balanced when it is rotated slowly and stops at any point readily.

Step 3 Number your weights with masking tape labels as 1, 2A, 2B, and P for plumb bob. Make sure that the holes are numbered. Arrange an experimental setup like that shown in Figure 2. Label the #1 and #2 weights with masking tape. By adjusting the paper harness triangle hanging from the pivot nail, you will be able to arrange to have the ruler touch lightly against the threads. You can adjust the height of the ruler by using the cuts in the paper harness triangle so as to adjust the thread to the necessary position. To keep the extra thread out of the way it may be laid on top of the triangle.

No cotton for damping the plumb line is necessary for when the apparatus is adjusted properly, the ruler will rest lightly against the lines and the slight friction will provide damping.

Once the apparatus is adjusted and hanging freely and has come to rest, you are ready to start the first part of your report.

Report - Section 1

Begin your report by copying the column headings of Table 1 below. Allow enough space for putting four lines of entries in the table. Write down the holes used for the weights. Measure and write in the Table the length of the two lever arms in cm., L_1 being for the single unit weight and L_2 for the double unit weight.

© Wm. Shockley 1964

019

TABLE 1

Holes used for

Weight 1	Weight 2 twice as heavy	length Lever 1 in paper spaces	Lever 2	2× lever 2 ~~$\frac{\text{Lever 1}}{2}$~~	$2L_2 - L_1$ L_2 ~~$- \frac{L_1}{2}$~~	$\left(\frac{2L_2 - L_1}{L_1}\right) \times 100$ Percent L_2 $\frac{L_1}{2}$	holes
							4-2
							2-4
							7-11
							1-11

The third column of the table is obtained by multiplying L_2 by 2.

In the next column of the table the difference is obtained by subtracting L_1 from $2L_2$. This should be written as a minus quantity if L_1 is bigger than $2L_2$.

The fifth column of the Table expresses the difference $2L_2 - L_1$ in terms of a percentage of L_1. It gives the percentage by which $2L_2$ is greater than L_1 as percent of L_1. If $2L_2$ is less than L_1, this percentage should be put down as minus.

Step 4 Repeat Step 3 using the three different combinations of holes for the spokes to sit in as follows: ~~1,2; 4,5; 1,5; 1,4; 2,5~~ see above Enter the new values of L_1 and L_2 in the Table.

If any one of your percentage is very different from the other three, repeat the measurements by putting the spokes back in the holes used before and making sure that there is no extra friction such as a plumb line pressing against the board or a weight line caught on a rough spot of the paper ruler.

Report Section 2

Why is L_1 not quite equal to $2L_2$? How do you expect it to differ from $2L_2$ and why?

What do you conclude about the ratio of L_1 divided by L_2?

Step 5 Use the holes that made L_1 and L_2 biggest in Table 1. Hang washers or paper clips on whichever weight is necessary so as to make L_1 equal to twice L_2; (you can probably be accurate to about 1 mm if you are careful.) It will be good enough if L_1 is within 1% of $2L_2$.

Report Section 3

Make a new table with the same column headings and call it Table 2. Write down L_1 and L_2 as you did for Table 1 but with the weights adjusted as for

© Dr. Shockley 1964

20

step 5 so that L_1 is almost exactly equal to $2L_2$.

Now move the spokes in the holes for three additional readings, with the weights adjusted as in Step 5. Enter these in Table 2.

Report Section 4

What do you now conclude from your experiment? Try to state a principle which applies.

Step 6 Get a third spoke and hang three nearly equal weights in three holes. Select the holes so that the weights and strings do not press against each other.

Report Section 5

Make a new Table 3 as shown below. Make sure your weights are labeled with scotch tape so that you can keep track of them.

TABLE 3

HOLES USED FOR						ALGEBRAIC
Weight 1	Weight 2	Weight 3	L_1	L_{2A}	L_{3B}	Sum of $L_1 + L_2 + L_3$
			~~10.3~~	~~3.2~~	~~14.1~~	~~0.6~~

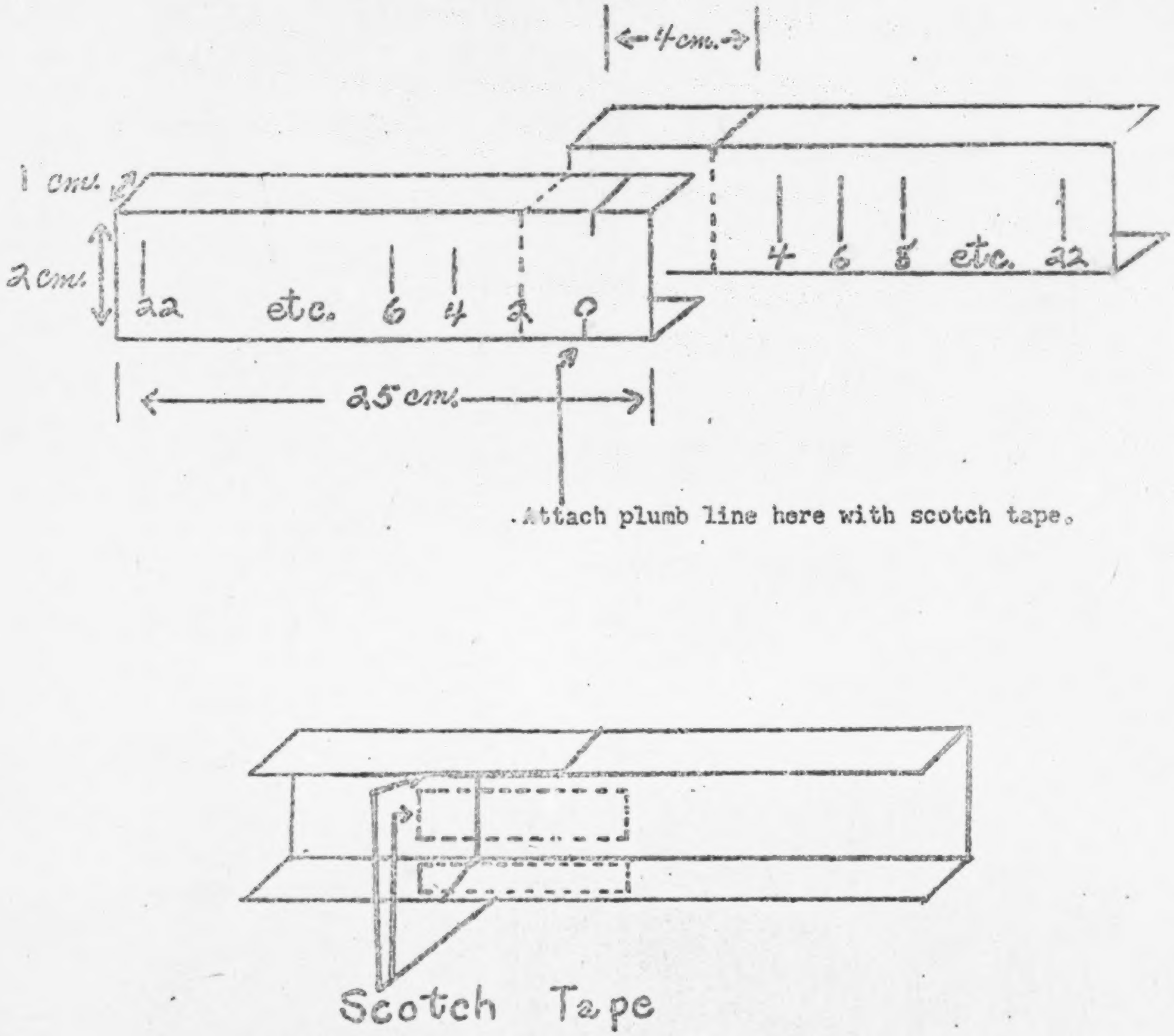
In this experiment consider any lever arm which has the weight to the left of the plumb line as negative and to the right as positive. (Notice that a positive lever arm hangs a weight on the board in a direction which tends to make the board turn clockwise.) Add up the sum of the lever arms and put this in the Table.

Step 7 (Report Section 5 continued) Repeat for two other combinations of holes and enter them in Table 3.

Report Section 6

What do you now conclude about the principle that must apply when a number of weights are hung from pivot holes on the board? Try to state this as a principle.

© Dr. Shockley 1964

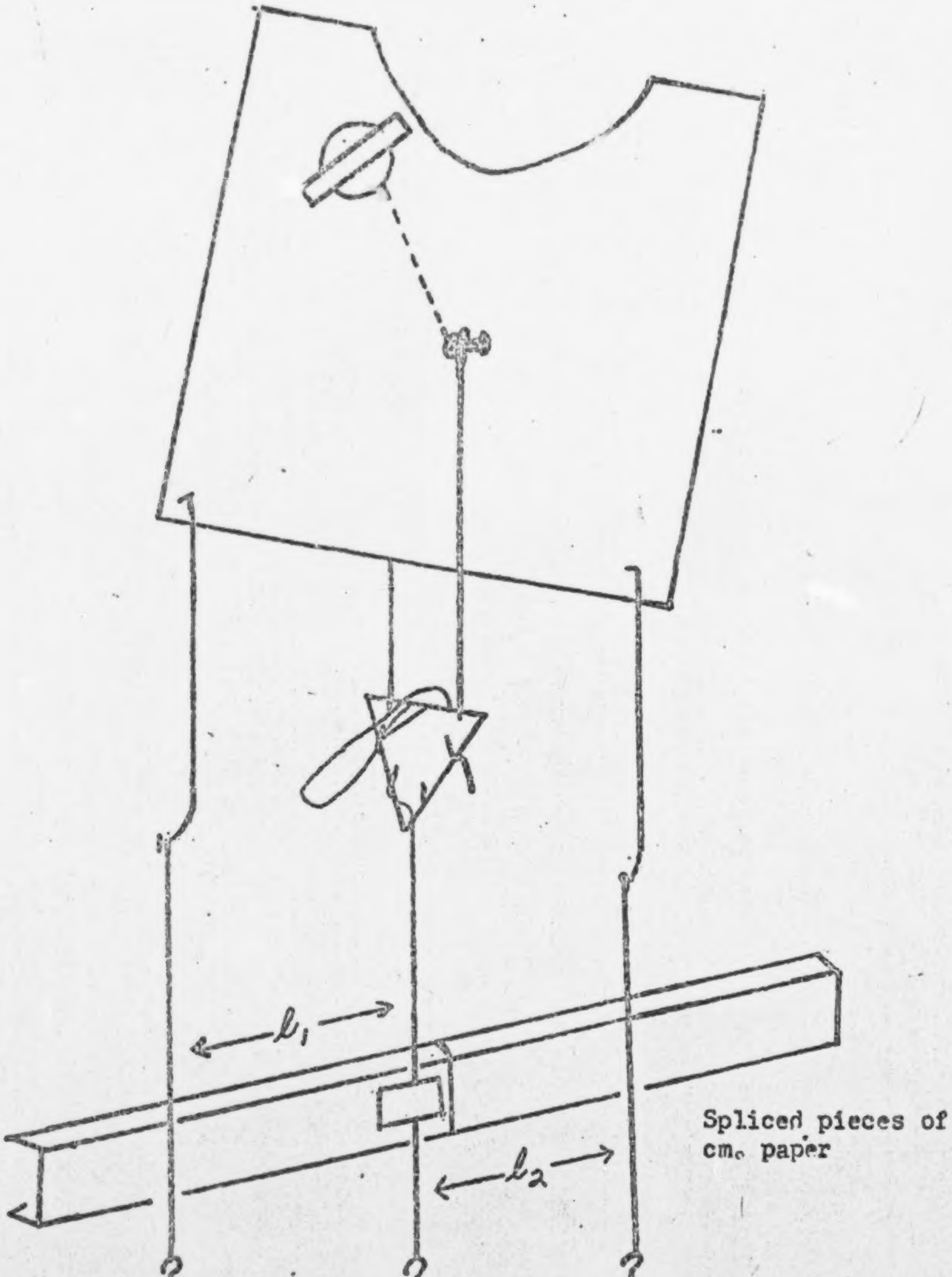

4 cm.
1 cm.
2 cm.
22
etc.
6
4
2
0
25 cm.
4
6
8
etc.
22
Scotch Tape

Attach plumb line here with scotch tape.

FIGURE 1 Making the lever arm ruler

© Dr. Shockley

FIGURE 2 Measuring the lever arms


l_1
l_2
Spliced pieces of
cm. paper
1
P
2A
2B

© Dr. Shockley. 1964

SUMMARY WORKSHEET

9TH GRADE SCIENCE

Name
Date
Period

Complete the following statements;

1. From the games of one and four, one through five and two pile, I learned --------

2. From working center of gravity experiments with the board, I learned that -----

3. As weights are added to a rigid object, the center of gravity---

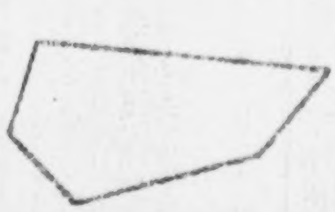
4. Two methods for finding the centerof gravity of a rigid object are;

   a)

   b)

5. If a 4000 pound piece of machinery anchored in the storage compartment of a flying airplane suddenly broke loose and started rolling,--------

6. The center of gravity of the figure below would be at --------



7. From the experiments with the board, having to do with the law of moments. I learned ---------

8. When two pounds is suspended from one end of a balanced lever and one pound is on the other end the ratio of the two lever arms must be--------- (Show with the aid of a diagram)

9. One source of error in them moments experiment was -----------

10. To the following diagram, add the fulcrum in the correct place and fill in the correct lever arms. (Add it on the back).

APPENDIX 2

TESTS

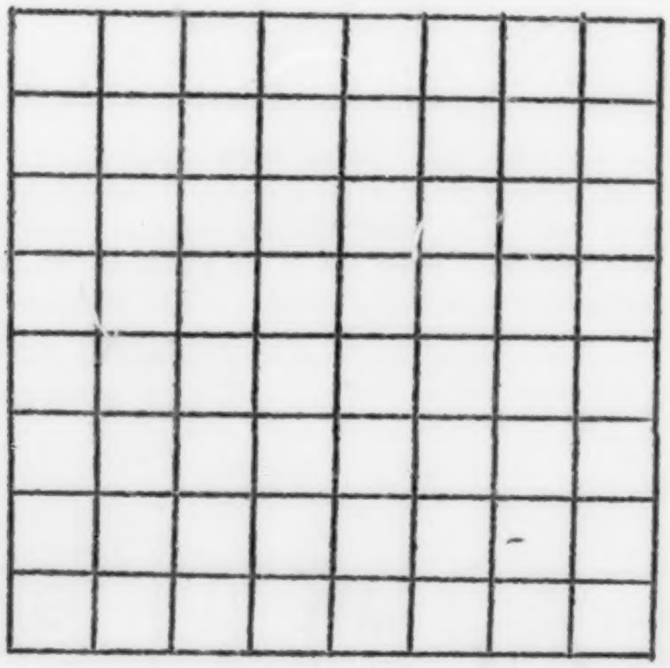
W. Shockley 18 February 1964

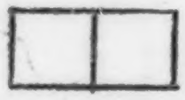
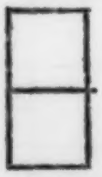
Checkerboard Test 1 (15 minutes)

"Graph paper" with about ¼" squares is needed. At the end of twelve minutes, teacher should ask students to write best answer they can to Problem 6 and then do what else they can on the test.

Checkerboard Problems

These problems are concerned with checkerboards and dominoes. For our purposes a regular sized checkerboard is a piece of paper with 64 squares, like this:



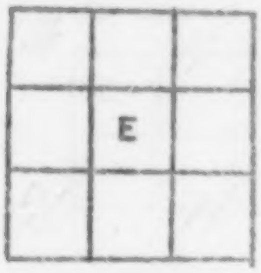
A domino is a piece of paper like this  which is just as big as two squares. A domino may be put on the checkerboard either horizontally or vertically .

You can use the answer paper, with squares to represent the checkerboard and short lines for the dominoes. The following example of a problem and your answer shows you how. To make it clearer, the problem is in a box and the answer is in another box.

© 1964 William Shockley

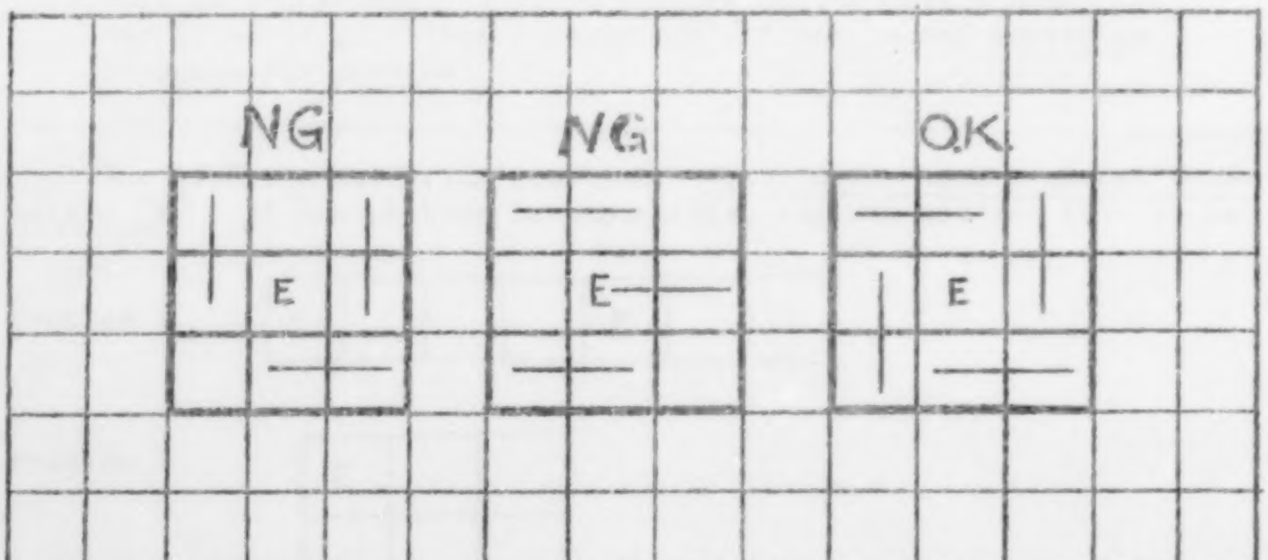
PROBLEM (EXAMPLE 1)

Put four dominoes on a three by three (3 x 3) checkerboard so as to leave the center square empty.



To answer this, take your paper and outline some 3 x 3 squares to work on. Represent a domino by a straight line from one square to its neighbor. Your answer should look like the answer example below, which shows two tries that were No Good and are marked "NG" and a correct answer marked "OK". Mark your unsuccessful tries "NG" in the same way and mark your solution "OK".

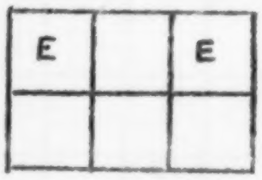
ANSWER (EXAMPLE 1)



Here is an example of the right way to answer an impossible problem:

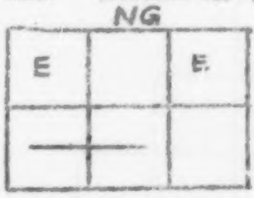
PROBLEM (EXAMPLE 2)

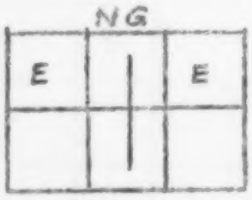
Put two dominoes onto a two by six checkerboard so as to leave the corner squares on the top side empty. These are marked E below:

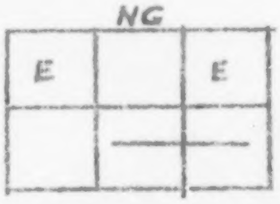


Your answer should show your work and state your conclusion and reasoning. The box below shows a good answer

ANSWER (EXAMPLE 2)



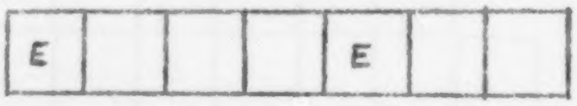




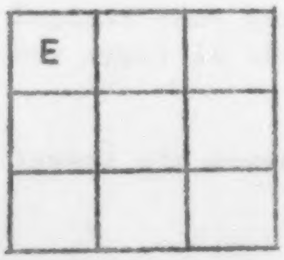
This is impossible. I tried putting on one domino every way it would go. Then I could not put the second domino on to solve the problem.

Now work the following problems. Cover every square except those marked "E". If the problem is impossible, explain how you know it is.

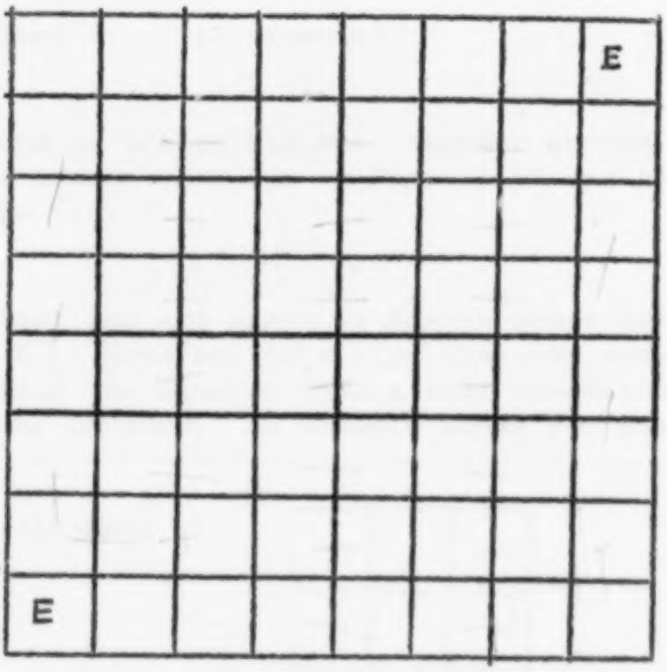
Problem 1



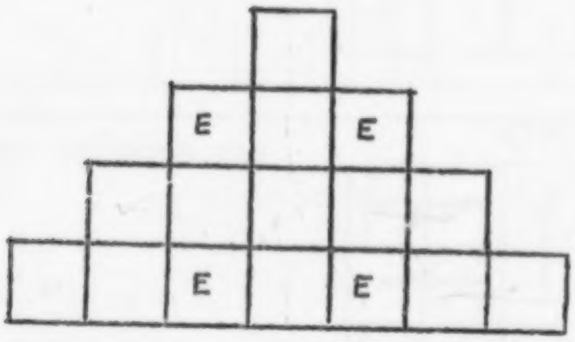
Problem 2



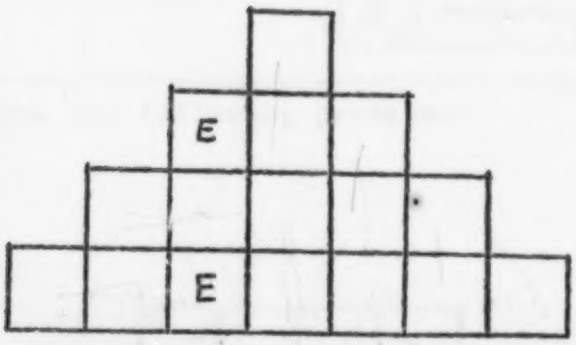
Problem 3



Problem 4



Problem 5



Problem 6 Explain what general principles or laws you believe may apply to these problems and say why.

Problem 7 Invent and answer problems of your own.

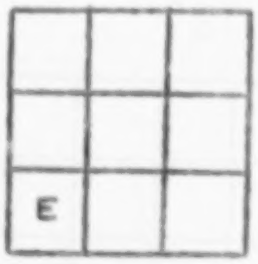
W. Shockley 18 February 1964

Checkerboard Test 2 (15 minutes)

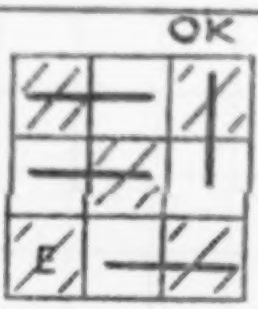
(At the end of twelve minutes, teacher should ask students to write the best answer they can to Problem 25 and then do what else they can on the test.)

In this test you are again to try to cover every square except those marked "E". However, this time when you mark out your checkerboard, shade half the squares like a real checkerboard before you try to work out your answers. An example shows you how:

PROBLEM (EXAMPLE 21)

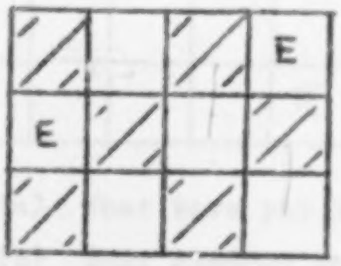


ANSWER (EXAMPLE 21)



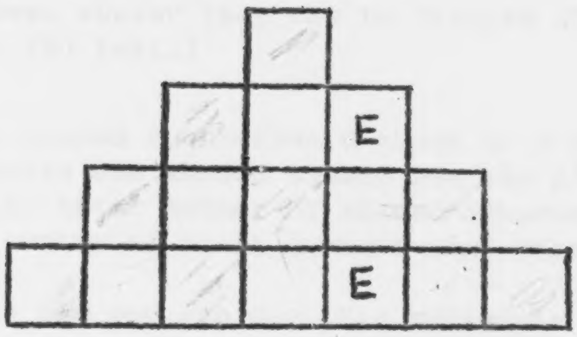
Now work the following problems:

Problem 21

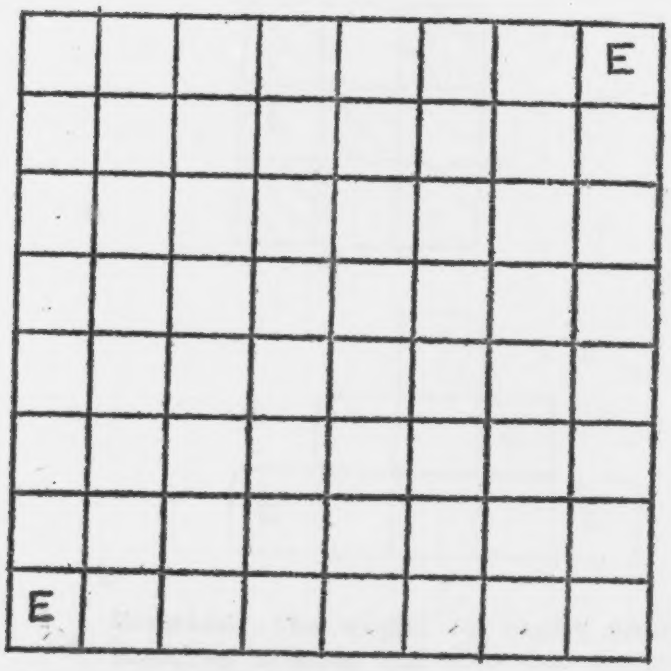


© 1964 William Shockley

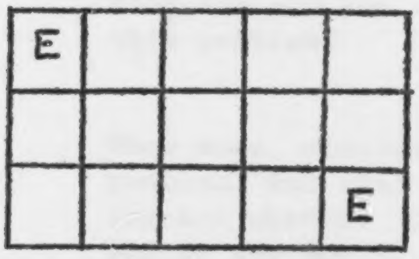
Problem 22



Problem 23



Problem 24



Problem 25 (a) What have you learned from these problems?

(b) What general principles can you use for analysis?

(c) State these principles as clearly as you can.

Problem 26 Invent some additional problems and their answers and discuss them.

W. Shockley 18 February 1964

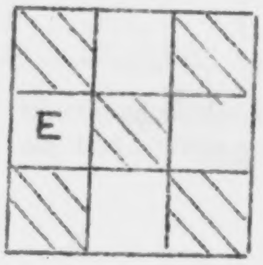
Checkerboard Test 3 (15 minutes)

(At the end of twelve minutes, teacher should ask students to write the best answer they can to Problem 37 and then do what else they can on the test.)

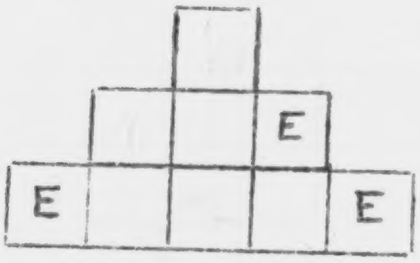
On the shaded checkerboard there is a principle. Each domino put down covers one shaded square and one plain square. It is always true that the total number of shaded squares that have been covered equals the number of plain squares that have been covered.

Explain how you can use this principle to understand these problems:

Problem 31



Problem 32



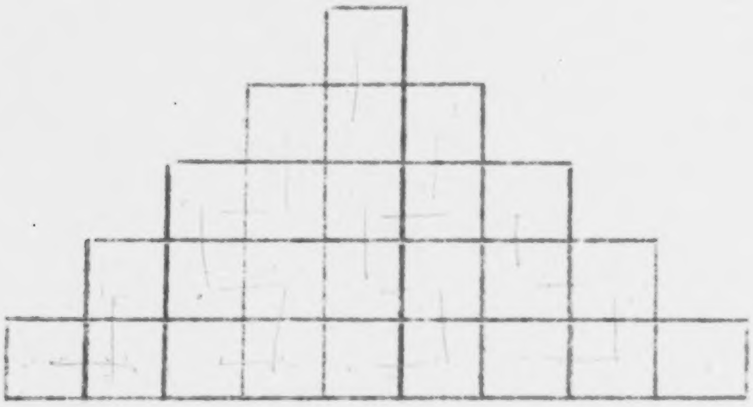
Problem 33 Consider the eight by eight checkerboard with alternate squares shaded and the two diagonal corners uncovered. What can you say now about the possibility of solving this problem?

Problem 34 Once more, consider the eight by eight checkerboard problem; but this time the squares on the checkerboard are not shaded. Can the problem be solved now? Explain why or why not.

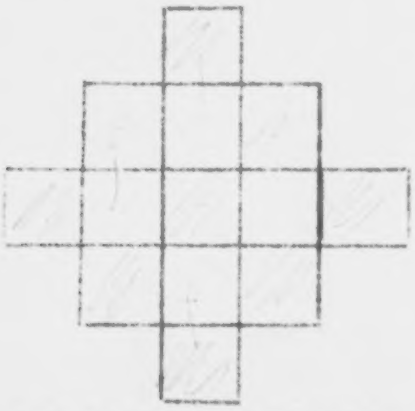
© 1964 William Shockley

Problem 35 What is the maximum number of dominoes you can put on this figure? Why?

25
15



Problem 36 What is the maximum number of dominoes you can put on this figure? Why?



Problem 37 Put in your own words in any form you prefer the principles you have learned to apply to these problems.

Problem 38 Invent and answer some additional problems and discuss the results.

30 Jul 64 W = S

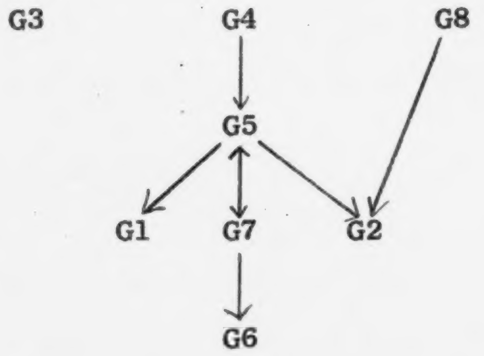
INSTRUCTIONS

Write your name at the top of each test sheet.

Question 1. In each of the lists put at the left of each statement a "P" if it is like a postulate and a "C" if it is like a conservation law. The others leave blank.

Question 2. List in order of usefulness in working problems the most useful statements in each list.

Question 3. Make a diagram like this for each list. This means G3, G4 and G8 are like postulates. G4 proves theorem G5, G5 and postulate G8 prove G2. G7 can be used to prove G6.



A. TAKE ONE OR FOUR GAME ("REMAINDER" IS AFTER ÷ 3)

A1. In the 20 match 3 player game, second player wins.

A2. A player always leaves the same remainder in a three person game.

A3. Each play reduces the remainder by one.

A4. The winner is the one who plays in the same position as the original remainder.

A5. Three plays bring the remainder back to the same value.

A6. The winner is the one who leaves a remainder of zero on any of his plays.

A7. The remainder is conserved by three plays.

B. MOMENTS, CENTER OF GRAVITY, POTENTIAL ENERGY

B1. A non-rigid body does not have a fixed center of gravity.

B2. A set of weights on a plane has a center of gravity.

B3. Potential Energy is conserved in a system with rotating balanced levers.

B4. For a balanced system the sum of the moments about the axis is zero.

B5. A rigid body will balance on its center of gravity.

B6. An ideal lever with weights attached so as to balance in all positions does not change its potential energy when it rotates.

C. CHECKERBOARD AND DOMINOES

C1. The problem of covering all squares can be solved only if the number of available squares is even.

C2. The difference between the number of uncovered shaded squares and the number of uncovered plain squares remains constant as dominoes are put down.

C3. The problem can be solved only if there are equal numbers of squares of the two kinds to be covered.

C4. Each domino covers one square of each kind.

C5. The number of covered squares is the same for each kind.

D. TENNIS TOURNAMENT (N = NO. OF PLAYERS AT START)

D1. Tournament is decided after N-1 matches.

D2. No. players remaining plus no. matches played is constant.

D3. To eliminate a player requires a match.

APPENDIX 3

SIGNIFICANCE TEST FOR THE RELATIVE RANK INDEX

APPENDIX 3. SIGNIFICANCE TEST FOR THE RELATIVE RANK INDEX

For the test of the null hypotheses in subsection 3D, each percentile interval in Table 10 is treated separately. In such an interval there are in general e students of the experimental group and c of the control with

$$n = e + c$$

These students actually have ranks in the entire population of the Table represented by

$$r_1, r_2, \ldots r_c, r_{c+1}, \ldots r_n \equiv S_r$$

where the first set of e ranks are those of the experimental group and the last c ranks belong to the controls.

The null hypothesis is that if the experiment were repeated again the same set S_r of ranks would be obtained by the students of the percentile but they would be distributed at random to the e experimental students and the c controls. The expected sum of ranks are then

$$\overline{r}_e = e\mu, \quad r_c = c\mu$$

where μ is the average rank

$$\mu = (r_1 + r_2 + \ldots + r_n)/n$$

for the percentile.

Actually the excess sum of ranks for the experimental groups is given by

$$RP_{EA} \equiv (r_1 + r_2 + \ldots + r_e) - e\mu$$

which is exactly opposite to that for the control group

$$RP_{EC} = (r_{e+1} + \ldots + r_n) - c\mu = - RP_{EA}$$

since it is easily seen that $RP_{EA} + R_{PEC}$ is zero.

The variance in RP_{EA} from its expected value of zero is a standard statistical problem. It corresponds to selecting at random e specimens from the set S_r. The variance is found to be

$$\text{variance } RP_{EA} = \text{average } (RP_{EA})^2$$

$$= \frac{r(n - r)}{(n - 1)} \sigma^2$$

where the variance of the set S_r is

$$\sigma^2 = \frac{\Sigma r_i^2}{n} - \mu^2$$

This quantity has been computed and tabulated for each percentile in Tables 10 and 11 on the basis of the actual ranks achieved.

On the null hypothesis, each percentile behaves independently. Consequently, the variance for the sum of excess rank points RP_{EA} for all the percentiles is simply the sum of the variances for all the percentiles.

The sum RP_{EA} for the whole table, being the result of many additive independent terms, should be well represented by a normal distribution. Consequently the one-tailed probability of a larger deviation than the observed RP_{EA} (obs) is given by using

$$Z = (\text{observed deviation})/(\text{standard deviation})$$

$$= (RP_{EA} \text{ summed over percentiles}) \div$$

$$(\text{variances summed over percentiles})^{1/2}$$

and finding the corresponding one-tailed probability in standard tables.

APPENDIX 4

AN EVALUATION OF DR. SHOCKLEY'S PROGRAM BY TWO TEACHERS AT WILBUR JUNIOR HIGH SCHOOL

An Evaluation of Dr. Shockley's Program by Two Teachers at Wilbur Junior High School

HISTORY

The experimental unit on "Learning To Think In Science About Energy" was officially begun on January 11, 1964. On that date Dr. William Shockley met with the three Wilbur teachers, Mrs. Dreyfuss, Mr. Bryner, and Mr. Donovan to lay the groundwork for what was planned to be a two week program at Wilbur, involving 6 ninth grade classes.

As the program developed, it was found that the original time estimate was grossly incorrect. The experimental unit lasted a total of 15 weeks.

Student activities and teaching materials were planned for on a week to week basis with teacher training meetings as the pilot sessions for designing these materials.

TEACHER PREPARATION

The teachers involved in the program met with Dr. Shockley for more than 30 hours as the unit was being taught, to receive instruction and to help develop the materials to be used.

As time went on, they took an increasingly more responsible role in developing teaching materials. The final month of the program was entirely prepared without the guidance of Dr. Shockley.

STUDENTS

Six 9th grade classes were involved in the program. Students were of average to high ability. These classes were organized as two team teaching units. It was, therefore, possible to have them meet in large group sessions of 80 students each. Dr. Shockley lectured to the large groups several times during the program. For the most part, however, classes were of typical size - about 25 students.

OBJECTIVES, METHODS, AND CONTENT

The teaching team had been trying to stress the method of inquiry - discovery approach - during the first semester of General Science. Dr. Shockley's program, therefore, fit this objective very nicely.

His aim was to guide the students into problem situations and show them how one arrives at broad general principles. The first problem situations were games. The students were able to discover the underlying reasons for winning each of the games. The thinking tools needed to solve the games were analyzed into steps. The use of one-to-one correspondences was stressed. Consideration was given to exceptions to rules. Emphasis was placed on how to state a principle or law. "It is always true that ______ no matter ______ but not if ______."

Then exact principles of nature concerning energy were "discovered" by the students individually as a result of carefully structured experiments. The physical principles included: center of gravity, law of levers, potential energy, the law of the pendulum - potential and kinetic energy, and conservation of energy.

The students were provided with written laboratory guides, with the materials needed for the experiments. There was minimal teacher guidance.

EVALUATION OF STUDENTS (BY TEACHERS)

Generally, at first, students were not favorable to the program. They were not sure what was expected of them. There was an insecurity instilled by the lack of a book to which one could go for ready answers. They were insecure about their own data or about observations they had made themselves. Frequently, teachers were asked: "Was this supposed to happen....?" "Is this the right answer?" They were also insecure because the teacher did not supply the ready answers, but parried questions with still other questions.

As time went on, there was an attitude change that was clearly discernible. The students began to trust their observations and data. At times, it was obvious that the students relished the challenge of a new problem. The complaints stopped and enthusiasm was detected.

Another general anecdotal statement could be made about the student who seemed to blossom under this type of learning organization. The student referred to is the one who is not challenged by the ordinary classroom situation where retention of ideas offered them by expository teaching is rewarded. Such a student showed a resourcefulness, improvising skill, and flexibility that the student who constantly was unsure about his answers (some of these latter students were top students!) never showed during the entire program.

The ability to apply mathematical skills was evidently lower in some instances than the program demanded. For example, we had to teach ratio and how to figure out percent error. Also, many students were not able to apply simple graphing techniques.

It was apparent to the teachers that the training in independent thinking was valued by the students, once they gained familiarity with the approach and the program.

EVALUATION BY TEACHERS

As a whole, we found the program very challenging and rewarding individually. We gained a great deal from the teacher training sessions with Dr. Shockley.

As we look back, we feel that the time involved in preparing us to teach was inadequate. Since the program was not in our hands at the outset, it took a long while for us to get "the whole picture". And, since preparation of materials was always on a hand-to-mouth basis, we were most insecure at times.

There were occasions when we seriously questioned the amount of time being devoted to this program. It is one of the criticisms leveled at the discovery approach that the time-cost involved is tremendous. Our judgment is that the program could certainly be shortened to about 6 weeks - and still retain all the very worthy features of training for independent thinking.

A high point in the program came when we prepared an unfamiliar problem for the students. This problem was based on previously arrived at generalizations. The students did remarkably well.

We recommend that the students themselves evaluate the program. We also recommend that before the program is again tested that the teachers involved be adequately trained prior to the start of the program and that the written materials be carefully reviewed before they are utilized.

APPENDIX 5

EVALUATION OF SCIENCE QUESTIONNAIRES AT WILBUR JUNIOR HIGH SCHOOL

PALO ALTO UNIFIED SCHOOL DISTRIC
ADMINISTRATION OFFICES • 25 CHURCHILL AVENUE • PALO ALTO, CALIFORNIA • DAVENPORT 4-4761

SEP 24 1964

September 23, 1964

Dr. William Shockley
Applied Electronics Laboratory
Stanford University
Stanford, Calif.

Dear Dr. Shockley:

Here is the list of questions used in the evaluation of the ninth grade science course by Mr. Donovan and Mrs. Dreyfous.

Part I

1. This science class as compared to other science classes you have been in.
2. Meeting occasionally for films, lectures, and tests in large groups in the cafetorium.
3. Laboratory experiences during the first semester.
4. How much I have learned this year in science.
5. Having more than one teacher in a team teaching arrangement.
6. Skill I have acquired in solving problems.
7. Amount of homework and study assigned.
8. Ability to follow the laboratory work sheets in discovering principles.
9. Rotation of classes during first semester for lecture and laboratory.
10. Overall value of the course.

Part II

1. My understanding of how a scientific principle is arrived at is:
2. The reason we learned games at the beginning of the unit was:
3. The discovery of principles by doing laboratory problems is:
4. The connection between the principles applied to energy and the games is:
5. As a result of the program, I feel that my skill is solving new problems has:
6. The reading material used was:
7. Dr. Shockley's lectures to us were:
8. The principles I learned by discovering them for myself are:
9. My confidence in my own ability to solve problems is:
10. My confidence in myself to use laboratory materials to perform experiments and collect and use data is:

Sincerely,

Nicholas Anastasiow

Nicholas J. Anastasiow
Director of Research

NJA:cw

SEP 2 1964

PALO ALTO UNIFIED SCHOOL DISTRICT
RESEARCH DEPARTMENT

September 1, 1964

To: Anne Dreyfous, Mike Donovan

From: Nicholas Anastasiow

Re: Evaluation of Science Questionnaires

The questionnaires you administered to your class have been analyzed and the results reported below.

You will recall that the first ten questions of the questionnaire deal with an evaluation of the course itself, while the last ten are related more directly to Dr. Shockley's program. I divided the responses into positive (Excellent, Good) and negative (Fair, Poor), and computed the differences from chance in a positive or negative direction.

The vast majority of both classes checked most of the first ten items in a positive direction. The chi square values of the chance of this occurrence are reported below. All of the p values in this study are results in a positive direction.

Analysis of Student Questionnaires
Ninth Grade Science Instruction

	Dreyfous 5		Donovan 5		Dreyfous 2		Donovan 2	
Item	chi sq.	p	chi sq.	p	chi sq.	p	chi sq.	p
1.	15.66	.001	35.35	.001	9.86	.001	57.86	.001
2.	38.10	.001	29.82	.001	18.64	.001	22.70	.001
3.	39.80	.001	24.93	.001	11.26	.001	[illegible]	.001
4.	18.84	.001	40.17	.001	16.75	.001	60.75	.001
5.	25.30	.001	22.73	.001	7.25	.001	1.20	N.S.
6.	4.68	<.05	5.34	.02	1.09	N.S.	1.39	N.S.
7.	23.05	.001	20.85	.001	2.52	N.S.	49.17	.001
8.	.70	N.S.	23.14	.001	3.31	.06	17.68	.001
9.	21.96	.001	8.70	.01	10.77	.01	14.42	.001
10.	32.68	.001	50.87	.001	29.02	.001	50.25	.001

There are some differences among the classes but these may be chance fluctuations. They are not, however, statistically significant, and it i recommended that each teacher interpret why this class and section responded in that way. For example, Class Dreyfous 5 responses were about equally positive and negative on question 8. Class Dreyfous 2 did not check items 6, 7, or 8 more frequently in a positive direction, nor did Class Donovan 2 check

5

items 6 and 6 more frequently in a positive direction. It appears that the skill they have acquired in problem-solving is not as definitely positive as most of the other items.

In analyzing the last ten questions, all sessions were comgined and the remarks children made were classified from positive to negative. Two raters independently scored the items (inter-rater correlation .90-.94). The chi square values reflect the difference this sample deviated from an expected 50/50 split. Students responses that answered the question rather than gave an evaluation were not included in the analysis.

Students responded significantly positively to items, "The reason we learned games . . ," "As a result of the program I feel that my skill in solving new problems has . . ," "My confidence to use laboratory materials to perform experiments and collect and use data is . . ." The students responded negatively to "My understanding of how a scientific principle is arrived at . . ," "The reading material used was . . ," "Dr. Shockley's lectures to us were . . ," "My confidence in my ability to solve problems is . . ."

There is some serious doubt in the minds of some of the pupils about the reading material, Dr. Shockley's lecture, and confidence in their ability to solve problems, more than half of all ability groups responding negatively to these questions.

Summary

It would appear that the methods Mrs. Dreyfous and Mr. Donovan used were stimulating and positively perceived by the students. The pupils, however, did not always perceive the value of Dr. Shockley's lectures. Many of the comments indicate that they were not always aware of the content or the context of his material and lectures.

NJA:cw
cc: Dr. Santee, Dr. Shockley, Dr. MacDonald, Dr. Koolen

APPENDIX 6

Report by Miss Humphrey on Cubberly
Teaching Experiments, Summer 1964

(A complete collection of relevant material
for the Cubberly experiment has been recorded
on 35 mm microfilm)

INTRODUCTION

Dr. William Shockley's ninth grade science unit, "The Magic of Science", was taught twice during July, 1964. This is the report of the teacher of this experimental project.

Three points were under particular study in this project. First, is the material suitable for students of senior high school age? Second, can a considerable part of the teaching involved be done by students? Third, will a student who has completed this course do significantly better in CHEM Study Chemistry than one who has not had this training? The answer to the first two questions appears to be yes, although experimental conditions differed from the usual school situation in several important respects. The third question awaits statistical study.

The classes involved in this study were taught at Cubberley High School in Palo Alto, where two classrooms were made available. Miss Melanie Humphrey was the principal teacher; Dr. Schockley, the originator of the program, also taught part of the time. The students were recruited from among Cubberley students signed up for chemistry in the school year 1964-65. They were paid by the hour for participating in the program. They were given no grade or other evaluation for their work in the course, and they were aware that it would not appear in their academic records. The course took a total of 40 hours each time, spread over two weeks. We met for four hours each morning, with one fifteen-minute break after two hours.

The students were divided into two groups, and the course was taught twice, as follows: Group A consisted of ten students with very high scores on both ability and achievement tests. This group was taught during Phase One, the first two weeks. In Phase Two, these ten students became the "teachers" and taught the same course to the 18 members of Group B. Group B represented roughly a cross-section of the ability of the students registering for chemistry.

Group A consisted of five boys and five girls, nine entering 11th grade and one about to begin 12th grade. All were 90th percentile or above in achievement or aptitude standardized test scores. None fell below 80th percentile in either test. On the Iowa Tests of Educational

Development "Comprehensive" score, they ranged from 84th to 99th percentile, with a median between 91 and 93. On the Differential Aptitude Test "V + N" score (intelligence), they ranged from 80th to 99th percentile, with a median of 90.

Group B, consisting of 14 boys and four girls, overlapped the first group considerably in scores. (Both groups were high, due to the "natural selection" of electing to take chemistry.)

On the ITED "Comprehensive" score, group B ranged from 64th to 99th percentile, with a median of 91. On the DAT "V + N" score, they had a median between 90 and 95, and ranged from 44th to 99th percentile. Ten of this group were entering the 11th grade and eight were between 11th and 12th grades.

The course material was planned for a ninth-grade general science course; thus, some of the science topics covered were not new to the students. The intent of the course, however, was not so much to impart knowledge about science as it was to train for problem solving. A game or puzzle approach was used to develop a pattern for thinking, then experiments were approached in much the same spirit: a problem was presented, which the students worked with until they developed one of their hunches into a principle. Students were encouraged to find logical relationships between the games and the problems in science, and to discover thinking tools which can help them attack problems. The course was particularly aimed at discovering the usefulness of certain very fundamental principles, such as the conservation of energy. Specific science topics covered were: center of gravity, moments and the law of levers, and energy.

My report will deal with four aspects of the program: the student-teach-student technique; the topics considered; the materials used; and the course objectives. For each I will describe the facts of the program, the students' reaction (based on conversation and on written, solicited comments), the teacher's reaction, and evaluation and suggestions.

THE STUDENT-TEACH-STUDENT TECHNIQUE

The second phase of this project was taught primarily by the students of Group A, under the supervision of the regular teacher. This technique was studied because it was felt that the faster students in most classes represent a pool of teaching talent which could be used to advantage. We wanted to see whether students could be used as teachers, and what strengths and weaknesses they might have.

Students who had just completed the course (Group A) were used as the teacher group. Each member of Group A was assigned two students from Group B, on a random alphabetical basis. (The remaining student from Group A was used as a supervisor to coordinate and control activities while the regular teacher was out of the room.) These groups of one "teacher" and two "students" formed the basic teaching unit for the second part of the course.

Teaching was done in one of four possible arrangements, all of which were used at various times. One was the "individual group," consisting of the "teacher" and two "students." This group was used especially for discussion of difficult points, for problem solving, and for remedial work. The most commonly used unit was the "teaching group," consisting of two "teachers" and four "students." This size group was found to be especially favorable for discussion, such as for writing principles of games or experiments. The "two groups" of eight "students" and four "teachers" were used for synthesizing varying points of view, for hammering out difficulties within a group, or for summarizing work done in smaller groups. The "large group" of all students was used for lectures, quizzes, instructions, and any occasions when the regular teacher wished to speak to the whole group.

The flexibility we enjoyed was partly the result of the fact that we had two normal-sized classrooms in which to work. The daily schedule was as follows: the Group B students were given a quiz first thing every morning. While they were taking it, the day's plans were discussed with the "teachers" of Group A in the other room. When the quiz was completed (and frequently corrected) and any announcements had been made, the class broke into teaching groups in the two rooms, and the "teachers" presented the work for the day, which was done in individual groups, teaching groups, or two groups, as called for. The regular teacher could circulate among the groups, observing or giving individual help to "students" or "teachers" as needed. A student supervisor oversaw the room when the teacher was not present. In addition to this formal, permanent arrangement, we also used some informal teaching of slower students by faster ones in both Group A and Group B. This teaching approximated the probable classroom situation more closely. For example, in Phase One about half of Group A proved able to deal with simultaneous equations easily, while the other half had difficulty. The slower students were given practice with "word problems" while the faster ones were set to solving a challenging problem in algebra. When both groups were through, the faster ones taught their problem to the slower ones. Both groups of students seemed to emerge from such sessions with an unusually firm grasp of the subject under discussion.

Considerable class time was spent in Phase One on actual teaching methods. There were discussions and demonstrations to prepare the students of Group A for their teaching assignments to come. Likewise, before the informal ability-grouping teaching arrangements were used, some instruction was given to the faster students on how to teach. The students usually taught well, showing a high degree of patience and usually leading their "pupils" to discover the desired points. There were, of course, a few who did too much of the work and thinking themselves, but for the most part the "teachers" were aware that the objective was to lead their "students" to make their own discoveries. Several commented, "We had to find it out ourselves, and so should they." Of course, there was some tendency for discussion to be sidetracked onto affairs of mutual interest to the "teacher" and "student." Discipline was occasionally a problem, usually when the regular teacher was not in the room. One boy (out of 18) occasionally refused to do as his "teacher" told him to.

Student reaction to the technique was generally enthusiastic. At the end of the first week, the students of Group B voted unanimously that they preferred being taught by students, to being taught by teachers. They explained that they were freer to talk, ask questions, and to say that they didn't understand.

At the end of the two-weeks course, most students in Group B were still in favor of the method and suggested its use in regular school. They commented again on the increased personal attention and freer give-and-take atmosphere. However, this time nearly all expressed concern over the amount of "goofing off" possible with student-teachers, and four of the 18 "pupils" opposed using the method in regular classes for this reason. The other opinions ranged from mildly in favor to very enthusiastic.

The "teacher" group, Group A, was also enthusiastic about the method after Phase Two had begun. They showed a good deal of apprehension when I told them half way through their course, Phase One, that they would teach it to the next group. The use of two-"teacher"-four-"student" groups was their suggestion; it seemed to reassure them to have another "teacher" to share the responsibility. But once the second phase had begun, none of them showed or expressed any anxiety about their role as teachers.

At the end of the first week, the "teachers" all said they thought the technique was good and was "one of the strong points of this course." They commented that pupils feel more at ease with other students; that more individual help is possible; that it "gets the students who are ahead of the others busy;" and that "there is a better chance for everyone to get involved and interested in the topics."

The "teachers" did describe some weak points for the method even at the half way point of the course. Chief among these was the possibility that the "teacher" might not know the material too well himself, and might get confused or teach something wrong. As one girl commented, "somehow the 'teachers' must be made to retain all that they were supposed to have learned." All members of Group A commented on the difficulty of maintaining discipline, especially when the regular teacher was out of the room or when the "pupil" and "teacher" were good friends. The general feeling seemed to be that the system worked well if the students involved were responsible, and that "the 'teachers' should have real teachers help supervise the whole thing."

At the end of the second week the "teachers" were less enthusiastic, although nine of the ten still thought the method could be used in high school for short periods or "for a change." They commented on the need for good understanding and responsibility on the "teacher's" part, and for an adult teacher to supervise discipline.

Interestingly, a few of the "pupil" group thought the method would be a waste of time for the faster students acting as teachers, but none of the teacher group felt this way. In fact, the fastest boy in Group A and one most susceptible to boredom in class specifically favored the technique as a way of involving "those students who are ahead of the others" in class work.

From the teacher's point of view, the method presents certain difficulties, but I feel that its advantages outweigh the problems. The most striking improvement over more conventional methods is the degree of participating and involvement evoked from the students. Because each "teacher" has only two or three "students," each student can and must become personally involved with the problem under discussion. It is much harder to be what one student called a "sitting sleeper" when you teacher is sitting three feet from you and has nearly no one else to watch. Simply because of the very small groups which become possible, the student can be forced into activity. And given this increased personal activity, problems and other topics seem to become more interesting. It is rather inspiring to see a roomful of 20 students, all working in threes on a difficult math problem with concentration and quite evident and vocal interest.

Another advantage of the small groups is the individual work which is possible. Misunderstandings or confusion can be discovered and corrected far more readily than in a large group. Additional questions or sidetracks can be followed up more readily by advanced students (within the limitations of the "teacher's" ability). The regular teacher is freed to work with students on an individual basis, which is a definite advantage of this method. To some extent, the student-teachers are merely a device for achieving the well-known advantages of small-group instruction.

Beyond this, there seem to be some advantages to using students as the small-group leaders. The "pupils" seem more willing to ask questions, to argue points, and to admit confusion or ignorance, when the teacher is another student. The student-teachers, having just worked through the same problems themselves and made the same mistakes, tend to be very understanding. Their patience on occasion was truly remarkable. The "teachers" themselves seemed to acquire a very firm grip on the material studied—probably partly because they were anxious to teach it well and partly because they studied it twice. The "pupils" also seemed to understand the material well, except in the areas where the "teacher's" knowledge was shaky.

The method poses some mechanical difficulties. It can be a headache to supervise the many groups—especially when two rooms are used at once, as in this program. It is difficult to meet with the teacher group separately to outline plans, but such meetings are necessary. The teacher group occasionally complained that they did not know where the class was going, even when each day began with a twenty-minute planning session. They also needed occasional help in recalling the material to be taught.

Discipline was sometimes a problem. "Teachers" were occasionally unwilling or unable to control their classmates, or forgot their role as teachers. The result was occasional rowdiness, more usually wandering from the subject. Much of the wandering and nearly all of the rowdiness occurred when I was absent from the room, and stopped when I returned or was sent for. This would be much less of a problem if the project were carried out in one large room. I did not feel that the sidetracking into small talk was a serious problem. Some student time was wasted in this way, but probably not as much as is spent in daydreaming or inattention in a larger class.

The success of the method may be dependent to a large extent on the personalities of the teacher group: on their maturity and responsibility, their understanding of the material, and their patience and ability to teach. I was greatly impressed on all these counts by the ten high school students used as teachers in this group. The students and the conditions may not have been typical; but at least I have observed that high school students can perform well as teachers. We spent some time before any teaching venture discussing the objectives to be sought and the methods to be used. I feel this was essential to the good teaching I saw going on during the project. Once they realized that the objective was to lead the students to discover things for themselves, not to tell them answers, the teachers followed that objective both conscientiously and ingeniously.

One ever-present problem is that of individual differences. With small groups proceeding independently, some naturally finished well ahead of others. There are two possibilities for these groups: one is to have extra or more advanced material available for them to study; the other is to use them to teach slower groups. Either or both may be used, but something must be available for them. The problem of individual differences, present in all classrooms, could not be ignored this summer. In a normal classroom, the "teachers" would probably be recruited from the first students to finish a task, so that the method itself provides a way of turning the problem to advantage. I did observe that the fastest students were not necessarily the best teachers; in fact, the most brilliant students sometimes tended to be impatient or to lecture when they became teachers. However, this was not difficult to correct, since I was circulating and made suggestions to teachers.

I noticed that the students became accustomed to nearly continual activity and that they then tended to be intolerant of lecture presentations. I thought I detected more than the usual amount of restiveness when the students were collected in a large group for lecture or discussion—though their participation was fairly good whenever a problem was offered. A few commented that lectures were "boring."

In general, I feel the method is very promising and should be tried out under classroom conditions. The summer program was unusual on several counts: class lasted four hours daily; the students were not taking any other courses; and no grades were given. The group involved was also rather select. But I feel the method could well be used on an ability-group basis, using those who finish first to assist those who have trouble. The teachers should have a good grasp of the subject and should be told what sort of teaching is desired. The regular teacher should move from group to group, making suggestions to the "teachers" or working with the "pupils." The ratio of two students per teacher may not be the best; teachers were sometimes inactive, and three or four students may be a more reasonable load than two.

TOPICS

The topics covered in both phases followed roughly the outline of the ninth grade science unit. The plans for Phase One were changed several times during the two-week period, to deal with needs we had just become aware of. Phase Two was based on the experience gained in Phase One and is probably a better arrangement.

In Phase Two, topics were taken up in this order:

1. Pretests: the three "checkerboard tests".
2. The penny-board balance game: play and derive a principle.
3. Take-one-or-four: play, derive principle, do reading.
4. Standard form for a principle: read and fill in worksheet.
5. Take-one-to-five and two-pile: play and derive principle. Reading on take-one-to-five.
6. Take-one-to-five, last match loses: play, derive principle, do reading.
7. Tennis match: analyze, derive principle. (Some groups also considered the match-squares puzzle and the number-of-acquaintances problem.)
8. Experiments 1 and 2: existence of a center of gravity for an object.
9. Experiment 3: center of gravity shifts for a non-rigid object.
10. Experiments 4 and 5: locating the center of gravity for a symmetrical body, and for a pair of weights.
11. Penny-board: moments: define and use to balance board.
12. Word-symbol worksheet: express a "word problem" in equations and solve them with algebra.
13. Derive the expression for the center of gravity (x_o, y_o) of weights on a coordinate system.
14. Prove that the gravitational moment of a board is the same as if all its weight were concentrated at the center of gravity. (The board-in-strips problem.)
15. Find the center of gravity of a bicycle spoke and weight it on a penny-board.
16. See-saw problem: derive the law of levers.
17. Board-moment experiment: confirm law of levers and redefine moment.
18. Summary: relate the concepts of center of gravity, moment, and balance.

19. Weights-on-shelves problem: derive conservation of potential energy.
20. Quiz and discussion on logical relationships.
21. Uses of conservation of potential energy: derive Archimedes' Principle and the dimensions of a hydraulic lift.
22. Energy interchange situations: discuss different forms of energy and minimization of potential energy.
23. Pendulum experiment.

The conservation-of-energy principle and its applications are meant to be the climax of the course. If more time had been available, we would have gone on to analyze the pendulum's motion and to relate it to energy. We would have derived some basic proportionalities involving kinetic energy. And we would have tackled at least one quite difficult problem to show the usefulness of conservation of energy. This would have called for at least one more day (four more hours).

The students were at first quite mystified by the course topics, especially the inclusion of games. Many of them commented towards the end of the course that they only saw the pattern or the intent of the course as it developed. One boy in particular felt this was good; "let them find out what's happening themselves as they go along." But more students felt that the relationship between the games and the experimental part of the course needed to be made much more explicit. Apparently this remained a mystery for some. Four out of 18 in Group B saw no similarity between the match games and the center of gravity problem when asked to compare them in a quiz. (The other 14 said the method or process of solving was similar for both—quite possibly following a broad hint in the preceding question.) Even one member of the teacher group described the topics of the course as a "weird combination" useful only for "getting practice in writing principles."

The students seemed to enjoy the topics taken up. They all liked the match games, and while a few said the rest of the course was boring, more said it was interesting or even "great!" Some students seemed to remember most of their ninth grade science, and for them a good deal of the work with centers of gravity and even energy was review. (One of these said the course was "Mickey Mouse"—ridiculously easy.) Others insisted the material was all new to them.

Some students, particularly members of Group A, made suggestions about the topics. One said more match games were needed, another said fewer; the majority seemed to feel that the time allotted to match games was about right. It was suggested that more correlation be made between the games and the experiments. Quite a few students commented favorably on the progression from simple to fairly complicated material. Several seemed to feel that two-pile and take-one-to-five were easier to understand than take-one-or-four and should precede it. One student opposed beginning with the penny-board game because it was more difficult than the others. Finally, several felt that the conservation principles should be stressed more, and that more time should be given to conservation of energy at the end. One boy suggested spending less time on centers of gravity to allow for more work with energy.

This series of topics seemed to me as teacher to be fairly well adapted to the course objectives. The fact that some of the students had studied the material before was a problem. In Group A, one boy had had considerable prior experience with match games and tended to blurt out the solution before the others were even familiar with the problem. The "checkerboard" problem had been encountered in previous schoolwork, though only one boy remembered it. More serious than these examples (undoubtedly rare cases) was the number of students to whom center of gravity was already a familiar concept. I feel this is likely to be encountered in any above average group which has had ninth-grade science. There are several possible solutions. One is to use the students who recall the material to teach it to those who do not and then to go on and study it in greater depth. Another would be to select some other topic for a high-school, college-preparatory group. (If the unit were part of a biology, chemistry, or geology course, the science material would be selected from the course work. Anything would do as long as it was susceptible to an experimental, discovery approach.) As a part of a general science, physical science, or physics course, the present material would be satisfactory.

This course seemed to be about the right mathematical level for 11th grade. As the course was taught, it utilized both algebra and geometry. The geometry was particularly involved in experiment 4 and in the examples for conservation of energy. Algebra formed a substantial portion of the course, including the "word problem" worksheet and the center-of-gravity theorem. Basic operations of algebra were involved in nearly all of the course. The course might be adapted for students without geometry, but algebra is either a prerequisite or a necessary inclusion in the unit.

Two topics in this unit are among the course objectives and should be expanded and emphasized. One is the "word-symbol worksheet" (topic 12), dealing with the translation of words into mathematical symbols. The other is the final section on the usefulness of the conservation of energy, topics 21-23. These should have more attention devoted to them—preferably by making the course longer than 40 hours; if necessary, by cutting out some of the other work. More expendable sections might be topic 7, the tennis-tournament problem; topic 10, experiments 4 and or 5 on locating the center of gravity for special cases; and topic 17, the board-moment experiment.

As the course was taught this summer, there was not sufficient correlation between the games and the other material. Students did appear to notice a relationship based on the thought process and the methods of problem-solving (the thinking tools). Many of them did not seem to feel that this relationship was very important; and very few felt there was any other correlation or similarity. There needs to be more done to make the games and the problems seem similar and to make the similarity seem important. Perhaps written material analyzing experiment 1 in a pattern similar to the analysis of take-one-or-four is needed. New games with a striking logical similarity to one of the topics could be included at the same point in the course as that topic. A new format for discussion of the similarities would also help. But if the games are worth including (and I think they are), they should be more integrated with the remainder of the course.

In general, I think the sequence of topics on pages A3.9-A3.10 is excellent. I observed a real flash of discovery or insight now and then during the course, especially in connection with the following topics: take-one-to-five; two-pile; the see-saw problem; and the weights-on-shelves problem. I like starting the course with the penny-board game; it cannot be fully understood at the beginning but will be returned to. Topic 20, the quiz and discussion of logical relationships, does not necessarily belong at that point; it was inserted there for reasons of convenience. The course should be extended slightly to add discussion of the pendulum experiment and further work with kinetic energy.

The question of the best grade level for this course remains open. Students suggested all levels from the 7th through 12th grades; tenth grade got the most votes. Several said it should be required for incoming tenth graders, as a part of the biology class. Others said it would fit well into 9th grade general science if the algebra could be simplified, or into 11th-12th grade science. A few suggested it should be encountered twice—once about 7th grade for the thought process, and again in high school for the science. Most students thought science was the best format for the unit, though a few suggested math classes.

I especially like the suggestion of 10th grade, if a suitable experimental topic could be found early in the biology course—perhaps an environmental study, where environmental balance could be considered in terms like conservation of oxygen-plus-carbon-dioxide. This level seems good because it is at an introductory point in science courses but suitably advanced in math. I would like to see this course tried in average-ability high school general science classes, to see whether it is suitable for all ability levels.

MATERIALS

The written materials used in this course consisted of textbook-type reading materials, directions for experiments, tests and quizzes, worksheets, and in Phase Two, instructions for the "teachers." (No homework was assigned for this summer project.) The sources of the materials included Dr. Shockley; the teachers of the 9th grade course; the regular teacher for the summer program; and outside sources such as Feynman.

Reading material consisted of readings by Dr. Shockley on three of the match games, and reading on work and energy by Feynman. The former seemed to serve well the purposes of making clear to each student what went on in the games and of displaying the thought process. Some students found them boring because they were "redundant" or "too repetitious." The materials were used after a game had been played and analyzed, so as to give the students a change to work through the entire process actively first. In that case, the materials seemed like a repetition of the same thing. They should be used—they serve a good purpose—but perhaps they could be made shorter, realizing that the students already know a good deal about the games.

The readings from Feynman were good and seemed to be interesting. They should always be followed by discussion of their contents, or by a worksheet of questions. Otherwise the material (especially the reading on work) seems not to relevant. Work (in the physical sense) should be discussed and related to conservation of energy by some problems on the maximum amount of work which can be gotten out of a system.

More reading materials are needed in some places. I believe Experiment One should be followed by a reading on the same pattern as the match game readings, so the similarities will be more evident. I would especially like to see the logical-relationship scheme made up into at least one reading, to follow a match game. Then the same sort of scheme could be used throughout the course in discussion.

There were eight written experiments (Experiment 6 was not used in Phase Two for lack of time). Experiments 1, 2, and 3 are to be used together, and all are concerned with developing and refining the principle about the existence of a center of gravity. These worked fairly well, and the students seemed to understand what was involved most of the time. Experiment 1 and its writeup took about 30 minutes; Experiment 2 took 10 minutes and could be done as a demonstration with no loss of effectiveness; Experiment 3 took about 15 minutes. Most of the students realized right away that the "magic pivot-line point" was the center of gravity, and many balanced the board on that point on the tip of one finger to prove it. Experiment One was followed by a group principle. The students kept attempting to go beyond the experiment and to state additional information about the center of gravity. It should be made clearer (though not stated outright) that what is wanted is a statement that objects have a center of gravity, operationally defined as the point which always hangs below the support. Otherwise high school students attempt to go far beyond their actual knowledge. Experiment 2 was good and seemed convincing. Experiment 3 presented an unexpected difficulty: several students had trouble realizing that the board with weights attached by rubber bands was non-rigid. I think they tended to confuse "rigid" with "solid." The experiment should be introduced by a definition of rigid and the question, "Is this board rigid or non-rigid?" The extra weights should be quite heavy and the rubber bands large, since the object approaches rigidity if the bands stretch tight.

Experiments 4 and 5 deal with locating the center of gravity for some special, simple cases. Experiment 4 needs to be rewritten; students did not know what was expected. Perhaps the concepts "symmetry about a point" and "axis of symmetry" need to be reviewed or introduced first. The experiment takes about 20 minutes, including the writeup, but discussion is somewhat unsatisfactory, due mainly to lack of understanding of symmetry. If this could be cleared up, the experiment would be valuable, because symmetrical two-dimensional objects are invariably the first "simplest cases" the students want to try. Suggesting this was the first evidence the students showed of transfer of learning from the match games; they suggested trying "a simple case, like we did with the matches." Experiment five is worthwhile if the cardboards are already made and available. It can be done in groups of up to 6, as a game.

The penny-board balancing operation is written up as a game rather than an experiment; the second time it is played the concept of "moment" is introduced by the teacher. Incidentally, the students should never be allowed to say "weight times distance" without stating what distance is meant; the frequently became confused over this. They must realize that a moment is always with respect to an axis.

Experiment 6 is designed to apply Experiment 5 and to introduce some interesting variations on the center-of-gravity theme. It was used in Phase One but not Phase Two. It seems generally worthwile but not indispensable.

A new experiment needs to be written to take the place of Topic 15. The object of this experiment is to demonstrate that the moment exerted on something by gravity is exactly the same as if all its weight were concentrated at its center of gravity. For this it was suggested that an irregular, transparent, flat object be used on the penny-board. The experimental directions would go something like this: find the center of gravity of the object, using the penny-board (they do not need to be given explicit instructions how to do this); "weigh" the object in units of one penny, by considering the object as if all its weight were concentrated at the center of gravity; change the position of the object, leaving the center of gravity at the same position, and see if the same "weight" is obtained; decide whether the assumption is a valid way of dealing with the object. The commentary should make clear that the weight is not actually concentrated at the center of gravity, but that the board acts as if it were. There was some confusion over this point.

The balanced-board moment experiment is a problem. It should be worthwhile in that it confirms the law of levers (already derived mathematically from the center-of-gravity theorem) and refines the definition of a moment. But it took four hours, and the students did not understand it. They frequently mentioned it as one of the things about the course which needed improvement. I suggest it be rewritten along the following lines. Part 1: Balance the board. Enough hints should be given that they will realize how to do it using the center of gravity. Part 2: Explain the setup: ruler, position of weights, etc. Ask them to predict what the ratio between the two sides of the "ruler" will be when one weight is twice the other. Then have them measure these distances in various positions on the board. They should record the holes used, the weights used, and the lengths measured, in a table with room for more columns. Part 3. Have them compute in another column the ratio of the longer to the shorter side in each case. Ask them to compare this to their prediction and to discuss any differences. If the percent error is desired, the method of computing it should be explained (not just a formula given), and they should be told to compute it for each ratio in still another column. I think it could well be left out; students seemed to find the experiment difficult enough without the added mystery of the final column in the table. Part 4: Ask what they might do to make their results more consistent with predictions. Suggest changing the weight until one arm is just twice the other and checking to see if this ratio remains constant in different positions. They should be cautioned to decide carefully which weight it is that needs to be added to. Part 5: Ask them to discuss their results. Remind them to state how "distance" is measured. Suggest they may want to define moments slightly differently as a result of this experiment. Part 6: If desired, use three equal weights at different hanging positions, but I think the experiment contains enough substance without this.

The pendulum experiment is good as written. The number of entries required in each table (at least each table after the first one) should be cut down or the experiment will take more time than it is worth.

Quizzes were used fairly frequently during this course, and worksheets to be filled out in class were found to be very useful. The standard-form worksheet and the work-symbol worksheet in particular are a vital part of the course. A worksheet should also be made up for topic 18, the summary of the center-of-gravity work. In the summer course, students were simply told to write three or four sentences (in four-student groups) relating the concepts of balance, moments, and center of gravity. The answers received were extremely varied and I did not feel the students had actually acquired the clear understanding which such a summary should produce. A worksheet should be developed, either a series of leading questions to be answered or a paragraph with blanks to be filled in (like a teaching program).

One of the big gaps in this course is the lack of tests and other measuring instruments. There are a few tests which present problems to be solved: the checkerboard tests and the dot-link test. There are none specifically designed around the objectives of this course. Some progress was made toward designing a test of understanding of logical relationships, and that would be very useful, especially if the same format were used as part of the course.

It is extremely difficult to test the ability to think scientifically. It should be possible, however, to design a test or tests of the ability and willingness to use the following: conservation principles; symbols and mathematics; and certain other specific thinking tools such as pencil and paper, thought experiments, simplest cases. If some problems could be designed whose solution depends on the use of one of these methods, they would serve as a measure of how well the course's objectives are being fulfilled. There is also a need for a unit test of material learned, if the course is to be used in school. This would test largely understanding and ability to use specific science concepts of the program.

In Phase Two, there was some use of mimeographed material to inform the "teachers" of the day's plans, or to review or present the topics to be considered during the day. This was appreciated by the "teachers" and lessened the need for supervision. I recommend it whenever student-teachers are used. The "teachers" should be warned that it is for their use only, not for the students—particularly when the material includes a statement of answers or principles desired.

I feel there is a need for one more type of written material before the project can get out of the experimental stage. That is a detailed and complete teachers' manual. I have observed that understanding of the course decreases with each step further removed from Dr. Shockley the person is. I feel that possibly only Dr. Shockley completely understands what the course is attempting to teach; next come the people who have learned it personally from him; third, those who learned from them; still less those who learned from the third group; and so on. Some degree of understanding seems to be lost at each step of the teaching. The difficulty arises especially in the area of the relationship between the games, the science material, and other areas outside the course, such as geometry. The less well a person understands this relationship, the less willing he is to become involved in a discussion of it, and the less he finds to say when the topic does arise. A teacher in this situation does not teach anything about the relationship even if he understands it to some extent himself. I found it impossible to get Group A to lead a discussion of the course objectives at any depth greater than the thinking tools. Indeed, I sometimes felt quite insecure attempting to lead such a discussion myself, and thus tended not to provoke such discussion. The major symptom of a teacher's lack of understanding on this point is a tendency to stick strictly to the facts, whether of games or of science, and to concentrate on the material under study without analyzing it or relating it to anything else. If this program is to teach more than just some concepts of science and some games, the deeper objectives must be communicated to the teacher as clearly as humanly possible.

The teacher's manual should contain a list of the objectives hoped for, a description of each, and three or four detailed examples to illustrate each objective. For instance, the statement "conservation principles are useful" makes no impression unless it is accompanied by something like this: the derivation of Archimedes' Principle; an analysis of the hydraulic lift or similar easy application; comparative treatments of the chain lift using conservation of energy one time, the law of levers the other; and a study of kinetic energy using conservation of energy and observed facts about pendulum motion. Similarly, the logical relationships between the various games and topics should be illustrated in detail, preferably by the use of some pictorial scheme or diagram.

Along with the objectives, the manual should contain as much advice as possible on ways of teaching the course. Of course, the teacher must realize that his own role is that of catalyst and guide, and that the work of learning is to be done mainly by the student. It should be clear that the student is not always to be forced to use exactly the words the teacher might use; there is some range of correct answers. Teaching methods which are known to be successful should be suggested. These would include:

1. Leading questions, designed to draw the student's attention to a key aspect of the problem without giving away what is wanted.
2. Insistence that the student express things in his own words. For instance: if a student is stuck, ask him to state the problem he

is working on or the question he is trying to answer. If something has been explained to him, ask him to explain it back to you—or to another student who still does not understand. If he has written an equation, particularly after struggling with it, ask him what it says in words, and whether that was what he wanted to say. If he loses sight of the whole problem or gets discouraged, ask him what stage of the thinking-process he is in, or what thinking tool might be helpful. When he says he understands a problem, have him explain it and then write a formal principle.

3. Pencil and paper, or chalk and blackboard. "Writing it down" has been one of the major keys to learning in this course. Some examples: A pattern to point the way to the solution of a problem can often be found if a student will write down everything he knows (observed or "given") in orderly form. Word problems *must* be preceded by writing definitions of all symbols to be used (e.g., "d = number of dimes," *not* "d = dimes"). Tables of simple numerical cases are very frequently helpful in word problems. For instance, a student might write "I have twice as many dimes as pennies" as "2p = d." The teacher would suggest entries for the first column of the table, and the student would fill in the second, as follows:

number of dimes	number of pennies
1	2
2	4
3	6
d	2d

Writing a "word equation" can also be helpful, as in the following example: I have twice as many pennies as dimes.

The number of pennies = twice the number of dimes.
p = 2 x d

My experience shows that a student should never be permitted to "do it in his head;" writing down his thoughts in orderly form is undeniably good for a student, especially for the students who are reluctant to do so. The same is true of a student who says, "I know it but I can't say it." He is not finished with the problem until he *can* say it and write it down in a way which satisfies him.

4. Student-teachers. Suggestions on how to use students as teachers can be found above.

The teacher's manual should also, of course, describe the materials needed throughout the course: center of gravity boards, penny-boards, weights, and so on. It should also give expected results and answers for the course work.

In general, I found the written materials to be excellent, and the students also expressed satisfaction with them. I do recommend a few revisions, as outlined above. I also strongly recommend that a teachers' manual be written before the course is tried on any large-scale basis, where personal briefing of the teachers by Dr. Shockley is not possible.

OBJECTIVES

There was some uncertainty in my mind about objectives during the course of this project. I now believe that the course is aimed primarily at four points: 1) the relationship between mathematics and reality; 2) the uses and usefulness of conservation principles; 3) similar logical structures in widely differing problems; and 4) the thought process and tools to aid thinking. Some of these tools are: simplest cases: pencil and paper; thought experiments (including limiting cases); and symbols (which must be defined, the problem translated from words, the equations manipulated, and the results translated back into words).

The students in the course were not aware of all of these objectives. The members of Group B, the student group, stated that they felt the object of the course was 1) to teach how to solve problems (many mentioned the thinking tools, especially simplest cases) and 2) to teach facts about center of gravity and energy. A few seemed to be groping towards a feeling that the match games and the experimental part of the course had similarities beyond the best method of attack, but no one actually formulated this idea. The "teachers" did not teach any of the goals except thinking tools explicitly. Dr. Shockley has spent an hour or two with this group on logical relationships but it seemed largely lost on them. Conservation principles seemed less important to them as conservation principles than as facts about energy. The section of the course dealing with math was not usually mentioned. I know that some of the "teachers" were very impressed with the importance of the conservation principles, and others with the math-and-everyday-life emphasis, but very little of this seemed to have impressed their "pupils."

The members of Group A seemed to have a better understanding of the objectives at the end of Phase One than did Group B at the end of Phase Two. In their "final," eight of them mentioned the thought process and thinking tools, five mentioned conservation laws, and two mentioned the connection between mathematics and reality. Again, a few seemed on the trail of relating the logical structures of different problems, but none was very close.

I feel the main reason Group B was not very aware of the goals of the course was that Group A was unwilling to teach them. When specifically asked to talk about the course objectives, the teacher group got as far as the thinking tools and then demanded, "what else is there to say?" I feel they were reluctant to discuss this partly because their own understanding of the other objectives was shaky and partly because they seemed to feel that the goals were either too unimportant or too obvious to mention. For instance, Dr. Shockley once pressed Group A during Phase One to see that two principles were similar in providing a short-cut to a solution, without needing details. When that point was finally made, two or three others said, "We all saw that, but we didn't think it was what you wanted." After that, they were willing to write such "obvious" things on tests, but not to point them out to their friends. Some of them even felt that "everyone already knew" about the thinking tools, though usually it was only the other objectives which were considered "obvious."

Thus the first group seemed somewhat aware of most of the objectives (though not the logical relationships), but not convinced that they were all important or worthwhile. Most of Group B seemed not even to be aware that any importance was attached to objectives 1, 2, and 3 (as on page A3.22). Understanding of the course seemed to be strongly boosted by discussion after the course was entirely finished. Two students were questioned on the value of the course a few days after it was completed. They showed a better understanding of all the principles than I had expected, and they commented afterwards that they had not understood the course very well until they were asked their opinions. The act of analyzing the value of the course seemed to make the students aware of a good deal they had not realized, even when they did all the discussing themselves with very little hinting from the teachers as to what was wanted. I strongly recommend discussion of the course a few days after its conclusion.

Of course, it may not be necessary for the students to recognize the goals explicitly, as long as the goals themselves are fulfilled. I think both groups were better able to relate mathematics to reality and to translate words into symbols. I think both groups were much more aware of the value of pencil and paper and of simplest cases. I think both groups were aware that conservation of energy is a more important concept than they had previously realized. Beyond that I do not think the course reached. I do not think most of the students saw any basic similarities between the various problems tackled, and I do not think they would look for a conservation principle if given a new problem to solve. As far as I can tell without measuring instruments, the fourth objective was achieved quite well, the first and second fairly well, and the third hardly at all.

An important part of teaching to these objectives is a clear understanding of them in the teacher's mind. I personally had only a vague idea of what was meant by the logical relationships between different problems until I saw the format developed on the next-to-last day of class. I am sure I was quite ineffective as far as teaching that objective went. Dr. Shockley talked to the class on this point but the students did not seem to grasp the importance of what he was saying. For this summer I think we failed to get across the third objective. However, I think the new vehicle for teaching it could prove to be quite useful, and if it was used another time, I think some progress toward this objective could be made.

Speaking for a moment as a student of this material, I personally learned a good deal which was new to me. In particular, the usefulness of the conservation of energy principle has never been evident to me in spite of four years' study of science. If only for that, I would consider the course worthwhile. Also, I have observed myself transferring this type of principle to other problems. For instance, a former student came to see me with a complex physics-geometry problem which I realized could be easily solved by a "conservation of string" principle. I am sure I would not have seen that simple solution before this summer. I have not yet observed myself using the logical relationships, and I am not sure I fully understand that point even now. But I have made progress with it since Dr. Shockley expressed the problem diagrammatically, and I think that is the method which will prove most useful in teaching.

For the course as offered this summer, I feel three objectives are appropriate: to teach the usefulness of conservation principles, the usefulness of mathematics as a way of describing reality, and the similarity of methods and tools effective in attacking different problems. With some development of new material, a fourth objective could be added: to display underlying logical similarity of widely differing problems. I can serve as a personal witness that at least one of these objectives can be taught so that it is transferred to other problems; though I can also witness that the teaching of them is quite difficult and may not succeed with everyone even under the best conditions.

RECOMMENDATIONS

1. Using students as teachers should be tried in regular classrooms and in many subjects. I predict it will be found a valuable technique. I feel it is most appropriate when a portion of the class has reached a point of good understanding and the remaining students need more time and individual instruction to reach that point. I would make the following specific suggestions:

   a. A "teacher-to-pupil" ratio of one-to-one for remedial work, and of one-to-two-or-three for introducing new material, is probably best.

   b. There should be a specific assignment of material to be covered and a worksheet or problem sheet to be done by each "pupil" and turned in.

   c. Extra or further work should be available, or other provision made for those who finish early. (Perhaps the worksheet should be too long to finish in the given time, and students should be told just to go as far as they can.)

   d. The arrangement should probably not be used for more than a day or two at a time. It could be repeated every few weeks or few months as needed.

   e. "Teachers" should be those who know the particular material under study best, not necessarily those whose overall performance or ability is highest.

   f. "Teachers" should be given prior instruction in how to teach the material: asking leading questions, not telling the answer, insisting on activity, making the pupil write things down, and so on. Some demonstration teaching followed by discussion gives very good results.

   g. "Teachers" should understand well just what material is to be covered, and how, in a given period. A mimeographed "lesson plan" can be helpful.

   h. I suspect a random assignment of "pupils" to "teachers" is best. However, free student choice and deliberate assignment by the regulating teacher should also be tried.

   i. The regular teacher should not leave the room while small-group student instruction is going on.

   j. Lots of scratch paper, or better yet, blackboard space, should be available. Movable desks and a large room are helpful.

2. I recommend that the sequence of topics on pages A3.9-A3.10 be followed as given, with the following changes:

   a. More work should be added at the end involving the applications of conservation of energy and possibly some conclusions from the pendulum experiment concerning kinetic energy.

b. Topic 20, the quiz and discussion of logical relationship, belongs elsewhere. The format of this quiz should be evolved into a framework for consideration of logical relationship, and this framework should be used throughout the course to discuss each game or topic.

c. The word-symbol worksheet, topic 12, should be expanded and emphasized.

d. Experiment 6 should be inserted at some point before topic 18.

e. The following topics may not be in the most advantageous possible position in the sequence: topic 1, the checkerboard tests; topic 12, the word-symbol worksheet; and topics 14 and 15, the effect of gravity on an object.

f. The objectives of the course should be pointed out explicitly wherever appropriate and related to the material at hand as much as possible.

3. The program should continue to be tried in school. I would particularly like to see:

a. This same program tried in a high school, average-ability physical science course;

b. A new, shorter experimental unit written around conservation principles in biology, to use in the 10th grade.

4. I recommend that the written materials used in this course be continued, with the following exceptions:

a. Readings on the match games, experiment 4, and the moment experiment should be rewritten as described above.

b. New reading material should be written to follow experiment 1; a new experiment should be written involving an irregular transparent object; and a worksheet should be developed to help the students summarize their center of gravity work.

c. The scheme used in the logical relationship quiz should be developed into at least one reading assignment and used as a format for discussion.

d. Tests should be developed of the broader objectives of the course and of the science material.

5. A teachers' manual should be written so as to give teachers of this program the clearest possible understanding of the course and as many suggestions as possible on how to teach it.

APPENDIX 7

OBSERVATION OF THE SHOCKLEY PROJECT FOR TEACHING SCIENTIFIC THINKING AT THE HIGH SCHOOL LEVEL, BY GEORGE MILLAR, VICE-PRINCIPAL, CUBBERLY HIGH SCHOOL

McLean rec'd
24 Aug 64

OBSERVATIONS OF THE SHOCKLEY PROJECT FOR TEACHING SCIENTIFIC THINKING AT THE HIGH SCHOOL LEVEL

The general objective of this experiment was to develop and evaluate teaching procedures directed toward the attainment of ability to discover and use general principals having logical content similar to the general laws of science.

In order to accomplish this end, more specific objectives had to be developed. As I understand the project, the specific objectives can be stated as follows: 1. To teach the relationship between math and reality; 2. To teach some to the uses and usefulness of conservation principals; 3. To recognize and make use of the similar logical structure in widely differing problems; 4. To learn the thought processes and tools used to aid thinking.

A series of games and experiments was developed which, it was hoped, would accomplish the specific goals of the experiment.

The experiment was carried on in two parts: the Spring experiment using the six Science classes at Wilbur Junior High School and the Summer experiment involving Senior High School students at Cubberley High School. The two experiments involved children of different ages, the class length varied in each of the two experiments, the students involved in the two experiments held different motives, different teachers were used, and the method of teaching varied in the two experiments. Therefore I will treat the two experiments separately in this report.

The Wilbur Experiment

This experiment involved a total of six classes in Science at the ninth grade level. Three teachers were involved. Each teacher was responsible for one class of average students and one class of above average students. The material of the experiment was presented to the classes involved over a period of fifteen weeks. During the teaching period, and at the conclusion of the experiment, tests were given to the experimen group as well as to the controls. The purpose of the tests was to measure the ability of the student to handle rather abstract word problems using the systems of thinking developed in the experimental material. In my laymans view of statistics, the results were hopeful but inconclusive.

From the comments of both the teachers and the students, however, I feel that learning did take place in the classes at Wilbur and that the specific objectives of the course were realized to some immeasurable (at this time) degree. I would guess that the testing instruments are not yet developed to the point where they can measure for the specific objectives.

The experiment uses the problem solving approach to the field of science. Much work has been done in this field in the last few years and much new material is coming out constantly. While the problem solving approach is not new in the field of science teaching, this particular experiment does add to the field some formal material designed particularly to teach the approach to problems in general rather than any specific problem. Specific problems are, of course, used, but they are not necessarily problems that fit neatly in one particular discipline. Conservation principals, for instance, can be carried to all fields of science and math.

The Wilbur experiment pointed out a number of problems involved in teaching science by the problem solving or "discovery" method. The major problem is probably teacher preparation. It was obvious from the comments of the teachers in the Wilbur experiment that they had a good deal of trouble handling the material themselves. Much study time was required of the teachers and much teaching time was involved on the part of Dr. Shockley. If this program is to be expanded, released time should be given to the teachers to study the material and to broaden their general knowledge of math and science so that they can "field" questions thrown at them by their students. The methods used in in this course are much more demanding of a teacher than the standard textbook teaching method because the teacher allowes the students to follow their own leads rather than dictating the solutions to problems. If students are allowed to do this, it is necessary that the teacher have an excellent command of the subject so that students can be properly directed in their thinking. The teacher must know enough about the subject so that he can keep the students from going off in unprofitable directions and suggest alternatives when dead ends are reached.

The teacher also must be familiar with the practical uses of the subjects under experimentation so as to give the material real meaning for the student.

A second problem concerns the time involved in the unit. The Wilbur experiment consumed a total of fifteen weeks, and although the material was not used every day in the classes, the experiment did carve a good deal of time from the normal instructional pattern. Material of the problem solving nature does appear to be time consuming, because so much labor is involved on the part of the student in finding answers rather than being told the answers. It is impossible to "cover as much ground" with this method as it is with the conventional classroom system. I would suspect, however, that what is covered is learned better and retained longer, therefore the time spent is justified. The problem arises in convincing teachers and curriculum directors, students and parents, who are adjusted to the "chapter a day" teaching system, that this system, with its apparent inefficiencies, is, in the long run, more effective. I don't think that the material developed to date can prove this point. I would suggest more follow up tests be given, using the subjects of the Wilbur experiment. If a device can be developed to test last years subjects during the next two years, and if this device can produce positive results, then the program will be easier to sell.

A third problem concerns the proper location of this material in the curriculum. Should this material be taught in Science? In math? Should it eventually be broadened out to become a course in itself? If it is to be used at the Junior

High School level, what grade level would be appropriate? Last year the material was presented in a Science course. I believe it might better be moved to the Mathematics Department. for two reasons. One, the teachers in math at the Junior High level would probably be able to handle the material better because of their greater working knowledge of mathematics. Two, while the material is essentially mathematical in nature, it involves concepts useful in science and in some cases has direct science orientation. There is a need in math courses to stress the relationship between math and science and this material is beautifully designed to do just that. The problem here is finding time in the tight math curriculum to include units of this nature in the standard course.

Because some knowledge of algebra is required in order to effectively work with the material developed to date, I don't think that the material can be presented earlier than the ninth grade.

Eventually a separate course in logic might be developed for the high school level, but much work must be done before a course suitable for a high school level will be ready.

I was able to contact a few of the Wilbur students involved in the Spring Experiment and got a broad range of responses for the eight students I was able to contact by phone. The students knew they were in an experimental situation during the spring, and therefore did not express surprise when called. Generally the responses were favorable, but whether the responses can be considered genuine or partially swayed by the "Hawthorne effect" I am not skilled enough to determine. The students generally liked the games, could see how their work with centers of gravity might be applicable, and were convinced that problem solving is hard work. I would suspect from their comments, that they have retained the approach to problem solving they learned during the experiment, but the only way to be sure of this is to test the group again this fall. Two of the students thought the system was too time consuming for what was gained and one of them claimed she never understood what was going on. Some insecurity was expressed over the apparent lack of participation by the teacher.

The Cubberley Experiment

The Cubberley experiment involved students of a different age level, different motivations were present, and the course was conducted in a different manner, therefore slightly different results were obtained. Much of what has been said before however, will still apply. Again the teacher felt a degree of insecurity in working with the material in the open ended situation implicit in this teaching method. In her report, Miss Humphries stressed the need for a teachers manual to help her with the material. This teacher, being bright, and having a good background, could probably benefit from a teachers manual and might not need any extensive in-service training, but I don't think the same can be assumed for all math and science teachers. This material puts too much responsibility on the teacher for us to assume that all or a majority of teachers would be able to become effective with the aid of the teachers manual alone.

Aside from the refinement of the material itself, the major difference between the Spring and Summer experiments was in the teaching system used. The summer experiment leaned heavily on the "student as teacher" while the Wilbur experiment did not use this system. In the summer experiment, about one third of the students were started early. These students worked through some of the games and experiments and when the rest of the students were called in , the original group became teachers and undertook to teach the material to the new students on a two student to one teacher basis. The comments on this method were favorable in the beginning but tended to get less so as the course progressed. The major reservation seemed to be the loss of time through "goofing off" when the classroom teacher was not in the room. The "student teach student" system is not new to formal education. This system is used in many classes at various times although it is seldom formalized to the degree used in the summer experiment. In science courses as well as in Social Studies courses it is common for a teacher to group students for specific purposes, usually to increase the probability of free exchange of ideas or the establishment of common goals. The system does lend itslf very well to the problem solving situation and therefore is a very valuable tool to use with the Shockley material. During the school year, where a class is confined to one room and there is no additional instructional time offered to the student-teacher, certain mechanical difficulties will arise, and it will probably not be possible to use the system to the extent that it was used this summer. I would suggest that the system be continued and used as much as possible in the future experiments. Even if the formal student-teacher arrangement is not always possible, the use of small student groups (three or four students to a group) should be encouraged. With material of this nature and children of this age, the lecture method should be used sparingly.

General observations and recommendations

While the methods of teaching used in this project were not new and the teaching of science for an understanding of basic concepts is becoming the popular approach, still there is great value in this project in that the material itself is largely new. As the problem solving approach grows in popularity and acceptance at the secondary level, the demand for new and useable material will be magnified. I feel that this experiment has contributed some significant new material to the field. This material must be polished before it reaches it fullest potential and testing instruments must be developed to test for the goals set by the experiment. Some headway has been made in polishing the material and new testing instruments are still being tried, refined and tried again. On the basis of my observations and talks with both teachers and students, I feel that the material does have real value and that the objectives of the experiment have been reached in part. These observations,however, are purely subjective on my part, but I feel very strongly that this approach and this material should be expanded----that the efforts expended to date have been very worthwhile, and that further work should done with this material.

I would suggest that during the coming year the material be offered at the high school level as well as at the Junior High. In order to avoid having the same students in both the junior high and the high school projects, I would suggest that the high school to be used does not draw on the same children as the junior high if the experiment is to continue after the 1964-1965 school year. If Wilbur students are to be used, then Gunn or Palo Alto high school should be used instead of Cubberley. Cubberley draws all its students from Wilbur and it will be very difficult to bypass the students who were exposed to the material at the junior high level if both Wilbur and Cubberley are used.

On the high school level, I would suggest, with two exceptions, the use of this material in the Mathematics department. If the Science department is used, we would have to draw on students from the General Science or Earth Science courses. I think that the present curriculum in Chemistry, Physics, and Biology is too tight to allow time for this experiment. The General Science and Earth Science courses draw on students of low ability for the most part and on those students who have had no background in Algebra or Geometry. If this course were used with this group, it would have to be rewritten at a lower level.

One exception to this would be the Geology course offered at Cubberley. This course is designed to fulfill college entrance laboratory science requirements and does contain students with at least an algebra background and most probably

Geometry also. In addition, the Geology curriculum is still, to a degree, more flexible than any of the other appropriate courses.

The so-called "tech-prep" classes at Cubberley offer real opportunity for the use of this material. This is a special curriculum designed for boys of above average ability but who have a record of low achievement. This is a two year program beginning in the junior year and is designed to turn out a product that can continue on to a four year college or to a two year technical course such as those offered at Foothill College. All the students in the tech-prep program have had Algebra and Geometry and all supposedly have an orientation toward science. The groups tend to be small (about twenty five students) and very cohesive (they have three or four hours a day together).

In summary, then, the methods are not original but the material is new and exciting. I feel that there is definitely something of value in the experiment and would urge that it be continued and refined.